1989

Isolation And Characterization Of Human Ribosomal Protein Complementary-dnas

John Bryan Lott

ABSTRACT

Ribosome biosynthesis serves as a model of organelle assembly and gene regulation in both prokaryotes and eukaryotes. In Escherichia coli the isolation and cloning of ribosomal RNA genes and ribosomal protein genes was a prerequisite to detailed investigations of expression of these genes. Research into ribosome biosynthesis in eukaryotes has lagged behind in part because of the lack of data on the structure of the genes and mRNAs for ribosomal proteins.

This thesis details experiments designed to identify genomic or cDNA clones that encode human ribosomal proteins. Two methods were employed to achieve the goal of obtaining full-length clones that would facilitate expression experiments in the future. The first, cross hybridization, used a ribosomal protein gene or cDNA from another species to probe a human genomic or cDNA library. In one case a fragment of yeast DNA containing the 3' end of one of the genes encoding yeast ribosomal protein S10 (the homolog of mammalian S6) was used to probe human genomic and cDNA libraries. This approach failed due to insufficient homology between the coding sequences of yeast ribosomal protein S10 and human ribosomal protein S6. In another set of cross hybridization experiments a rat ribosomal protein S11 cloned cDNA was used successfully to

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISOLATION AND CHARACTERIZATION OF HUMAN RIBOSOMAL PROTEIN cDNAs

by

John Bryan Lott

Department of Biochemistry

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

February, 1989

ISBN 0-315-49306-2

Canada

ABSTRACT

Ribosome biosynthesis serves as a model of organelle assembly and gene regulation in both prokaryotes and eukaryotes. In Escherichia coli the isolation and cloning of ribosomal RNA genes and ribosomal protein genes was a prerequisite to detailed investigations of expression of these genes. Research into ribosome biosynthesis in eukaryotes has lagged behind in part because of the lack of data on the structure of the genes and mRNAs for ribosomal proteins.

This thesis details experiments designed to identify genomic or cDNA clones that encode human ribosomal proteins. Two methods were employed to achieve the goal of obtaining full-length clones that would facilitate expression experiments in the future. The first, cross hybridization, used a ribosomal protein gene or cDNA from another species to probe a human genomic or cDNA library. In one case a fragment of yeast DNA containing the 3' end of one of the genes encoding yeast ribosomal protein S10 (the homolog of mammalian S6) was used to probe human genomic and cDNA libraries. This approach failed due to insufficient homology between the coding sequences of yeast ribosomal protein S10 and human ribosomal protein S6. In another set of cross hybridization experiments a rat ribosomal protein S11 cloned cDNA was used successfully to

probe a human cDNA library and identify a full-length human cloned S11 cDNA.

The second method of isolation employed mixed sequence oligodeoxynucleotide probes to identify human ribosomal protein S6 cDNAs. Ribosomal protein S6 is the major substrate of protein kinases in eukaryotic ribosomes. Published amino acid sequence data for rat liver ribosomal protein S6 and yeast ribosomal protein S10 were used to design mixed oligodeoxynucleotide probes. Screening of several human cDNA libraries with these probes permitted the isolation of recombinant plasmids whose cDNA inserts encompass the entire coding sequence of S6 (249 aa residues), 27 bp of the 5' untranslated leader and all 39 bp of the 3' untranslated region. A comparison of the predicted amino acid sequence and the yeast rp S10 amino acid sequence shows highly conserved areas separated by regions of divergence.

ACKNOWLEDGEMENTS

I would like to thank Dr. Paul Berg for his gift of a human fibroblast cDNA library, Dr. Jack Riordan for providing an aliquot of a human placental cDNA library and Dr. Kikuo Ogata for providing pRS-11-2. I would also like to express my appreciation to my thesis supervisor Dr. George Mackie and to my fellow graduate students B. Cameron Donly and Philip [illegible] for advice, assistance and friendship. Special thanks to Marion Lott for enduring the frustrations of research and for typing portions of this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

Ab	- antibody
bp	- base pair
C	- degrees centigrade
C terminus	- carboxy terminus
CFU	- colony forming unit
cpm	- counts per minute
d	- daltons
h	- hour
kb	- kilobase pairs
kd	- kilodaltons
M	- moles
min	- minute
mg	- milligram
N terminus	- amino terminus
nt	- nucleotide
PFU	- plaque forming units
PIPES	- piperazine-N,N'-bis(2-ethanesulfonic acid)
uCi	- microcurie
ug	- microgram
ul	microlitre
rp	- ribosomal protein
SDS	- sodium dodecyl sulfate
SSC	- 0.15M sodium chloride-0.015 sodium citrate pH 7.6
X	- chi

CHAPTER 1

GENERAL INTRODUCTION

Chapter 1

a. Introduction

The ribosome functions as a key part of the translational machinery of the cell, but also serves as a model of organelle assembly and gene regulation. The interest in ribosomal synthesis centres on the mechanisms used by the cell to ensure an adequate supply of each component of the ribosome without needless accumulation of any single macromolecule. These mechanisms must allow the cell to respond to altered conditions of growth as well as changes in cellular metabolism elicited by some growth factors.

b. Structure of Eukaryotic Ribosomes

The eukaryotic cytoplasmic ribosome is composed of four RNA molecules and about eighty proteins divided between two subunits (Wool, 1979). The sedimentation coefficient for the entire rat liver ribosome, as measured in Svedberg units, is approximately 80S, while the subunits, known as 40S and 60S, are actually 36.9S and 56.4S respectively (Wool, 1979).

The small subunit, 40S, consists of one RNA molecule and up to thirty-three proteins. The apparent average

molecular weight of rat liver ribosomal proteins (rp), measured by SDS polyacrylamide gel electrophoresis. in the 40S subunit is 24,000. The 18S RNA component of the 40S subunit derives its name from its sedimentation coefficient. In humans, this molecule is 1870 residues in length (Gonzalez and Schmickel, 1986).

The large 60S subunit consists of three RNA species and up to 49 proteins (Wool, 1979). Based on their mobilities in SDS polyacrylamide gels, the ribosomal proteins have an average molecular weight of 21,000. The RNA molecules are known as 28S, 5.8S and 5S and in human they are 5025 residues (Gonzalez et al., 1985), 158 residues (Nazar et al., 1975) and 121 residues (Forget and Weissman, 1967) respectively.

Most ribosomal proteins are extremely basic with many exhibiting isoelectric points above pH 11. With the exception of E. coli S10 (nusE) which appears to function as an accessory to transcriptional termination (Friedman et al., 1987), no free ribosomal protein from any organism displays any enzymatic activity unrelated to protein synthesis. In other words, ribosomal proteins appear to function almost exclusively as components of the ribosome. Column chromatography and, more recently, HPLC have been used to separate ribosomal proteins from rat liver and other eukaryotic sources in amounts sufficient for amino acid analysis and partial sequencing

(Wool, 1979).

c. Comparative Investigations of Ribosomal Proteins

Since prokaryotic and eukaryotic ribosomes perform essentially the same functions many efforts have been made to determine if the ribosomal proteins from these different organisms share structural similarities. Initially comparative studies of ribosome structure relied on immunological or electrophoretic comparisons of ribosomal proteins of different species. More recent research has employed amino acid sequencing of purified proteins and DNA sequencing of cloned cDNAs encoding ribosomal proteins.

In an early study, polyclonal antibodies (Abs) directed against the ribosomal proteins of Escherichia coli were tested against the ribosomal proteins of other Enterobacteriaceae and detected few similarities (Geisser et al., 1973a). However, when these same Abs were allowed to react with ribosomal proteins from Bacillus species some immunological relatedness was detected (Geisser et al., 1973b). Interestingly, the ribosomal proteins of E. coli and the other Enterobacteriacae displayed similarities in their mobilities on two dimensional polyacrylamide gels, whereas E. coli and Bacillus ribosomal proteins exhibited great heterogeneity. The

authors concluded that the ribosomal proteins of E. coli and Bacillus are in part immunologically related but overall are structurally distinct.

Another study compared the ribosomal proteins of various eukaryotes by immunological methods. Rabbits were injected with ribosomal proteins from rat, chicken or rabbit to generate polyclonal Abs directed against the proteins (Delaunay and Shapira, 1974). Rabbit ribosomal proteins were not autoimmunogenic but some antibodies were raised against rat and chicken ribosomal proteins implying that these possess some epitopic differences when compared with rabbit ribosomal proteins. The anti rat ribosomal protein antibodies did not cross react with chicken ribosomal proteins and vice versa. The results led the authors to conclude that the antibodies were directed against some species differences in ribosomal proteins but that many ribosomal protein epitopes from rat and chicken were similar to ones in rabbit and were not immunogenic. This negative result was offered as proof of the high conservation of ribosomal protein structure in higher eukaryotes. Other groups have shown that anti-ribosomal protein Abs from one species can inhibit protein synthesis in another species (Fischer et al., 1978; Tanaka et al., 1980). The conclusion drawn from such experiments is that the Abs must recognize an epitope on the target ribosome sufficiently to block

translation of mRNA. Thus, while species differences do exist in ribosome structure, there are areas of high conservation that give rise to common determinants. Overall, the immunological approaches provided interesting yet inconclusive data and few dividends were realized.

Recently, the availability of amino acid sequence data, obtained by direct amino acid sequencing or inferred from DNA sequences encoding ribosomal proteins, has allowed more detailed comparisons of primary sequence to locate areas of conservation of structure between species. As an example, such a comparison has been made between yeast rp YP55 and rat liver rp L37. Although yeast rp YP55 is comprised of only 88 amino acids (Itoh et al., 1980) whereas rat rp L37 contains 111 amino acids (Lin et al., 1983), the N terminal sequences of these two proteins show striking similarities. Analysis using a computer alignment program revealed that of a total of 87 possible matches there were 47 points of identity between the two sequences.

Separated even further on the evolutionary scale are yeast and human and yet the amino acid sequence of an as yet undesignated human ribosomal protein, as predicted by a cloned cDNA sequence (Davies et al., 1986), shares 72% homology with yeast rp L44 (Itoh and Wittmann-Liebold, 1978). However, when the human cDNA clone was used to

probe Northern transfers of RNA from rat, guinea pig and yeast only rat and guinea pig RNAs displayed homologous sequences. Presumably, the functions of these proteins are sufficiently alike so as to require some significant structural conservation that can nonetheless be achieved without maintaining detectable nucleotide sequence homology.

Another study compared two of the phenotypic markers found in eukaryotic ribosomes. These were human rp S14 which confers emetine resistance (Chen et al., 1986) and yeast rp 59 which confers cryptopleurine resistance (Larkin and Woolford, 1983). Out of 151 amino acids, human rp S14 and yeast rp 59 are matched in 109 positions, that is, they have 72% identity in their amino acid sequences. At the nucleotide level, within the coding regions, there is 65% homology. In this same study human rp S17 and yeast rp 51 were found to share 77% amino acid identity in the first 91 residues with most mismatches being conservative changes. Further analysis of the nucleotide sequences encoding these proteins found that in their region of greatest homology human rp S14 and yeast rp 59 share 66 of 73 nt for 90% identity. Human rp S17 and yeast rp 51 share 36 of 44 for 82% identity. When DNA fragments from the coding regions of the human sequences were radiolabelled and used to probe Southern transfers of yeast DNA only the rp S14 clone annealed to

yeast DNA. One practical implication of this result is that the degree of identity, length of greatest identity and base composition are all crucial in determining the utility of probing DNA libraries of one species with DNA from another species.

It is not surprising that DNA sequences from different species that encode analogous proteins are more likely to be similar as the species are more closely related. Along with the obvious homology in amino acid sequence, the bias in codon usage appears to be similar in related species. For example, during the study of a human HLA-DP Beta 1 chain gene, Young and Trowsdale (1985) discovered a pseudogene in one of its introns. This pseudogene was used to isolate a cDNA clone homologous to the cDNA of mouse rp L32 (Dudov and Perry, 1984) from a human library. In fact the two coding regions predicted the same amino acid sequence and differed by only 36 silent changes although 170 changes were possible.

To investigate further the degree of homology between DNA sequences from different species encoding analogous ribosomal proteins, Meyuhas (1985) designed an experiment to determine the limits of cross hybridization of mouse ribosomal protein cDNAs with RNA from different species. RNA from rat, a human epidermoid carcinoma cell line, Xenopus laevis, Drosophila melanogaster, and yeast

were fractionated electrophoretically and Northern transfers probed with cDNA clones representing mouse rp L7, rp L18, rp L19 and rp S16. All four probes annealed to RNA from rat and human epidermoid carcinoma. In addition, the probes from rp L18 and rp L19 annealed to homologous sequences in the RNA from Xenopus laevis and Drosophila melanogaster whereas the rp L7 probe only annealed to Drosophila melanogaster RNA. The cDNA probe for mouse rp S16 did not anneal to RNA from the latter two species and none of the probes annealed to yeast RNA. Thus, whether or not it is due to codon usage or to the degree of conservation of amino acid sequence or both, not all mouse ribosomal protein cDNAs anneal to the same spectrum of homologs across species barriers. The degeneracy of the genetic code can account for the lack of cross hybridization between DNA sequences that encode proteins having similar amino acid sequences.

d. Functional Studies of rpS14 and rpS6

Individual ribosomal proteins may be involved in structural capacities, initiation, elongation or release of mRNA during translation or in interactions between subunits. Two ribosomal proteins in higher eukaryotes have received more attention than others since observable functions are associated with them. One of these is rp

S14 since it represents a phenotypic marker in conferring emetine resistance on the cell and the other is rp S6, the major phosphorylated rp of the eukaryotic ribosome.

Emetine is an alkaloid that prevents the puromycin induced breakdown of polyribosomes. That is, it inhibits the movement of ribosomes along the mRNA. Investigations of emetine-resistant mutants by Gupta and Siminovitch (1977) determined that resistance could be localized to the 40S subunit. The emetine resistance mutation was then shown to be associated with the structural gene for rp S14 in CHO cells (Nakamichi et al., 1986) which was used in turn to isolate the human rp S14 gene (Rhoads et al., 1986). The mutation that causes the phenotypic change occurs when two arginine residues near the C terminus are replaced by cysteine or histidine (Rhoads and Roufa, 1985). Roufa and his colleagues have been able to map the functional gene to chromosome 5q (Rhoads et al., 1986).

Another eukaryotic ribosomal protein that has received much attention is rp S6. Several proteins in the eukaryotic ribosome can be phosphorylated in vitro on serine residues and to a lesser degree on threonine residues (Kabat, 1971; Eil and Wool, 1973). Phosphorylation of proteins in the nucleolus has also been observed (Grummt and Grummt, 1974) and may represent

the phosphorylation of pre-ribosome associated proteins or rp on the precursor ribosome (Ballal et al., 1975).

In the cytosol, the major target of phosphorylation in mammalian ribosomes is rp S6 (Wool, 1979). Phosphorylation occurs mainly on a cluster of serine residues near the C terminus of the protein when cells are stimulated by a variety of agents. Serum, epidermal growth factor (EGF), prostaglandin F2 alpha (PGF2 alpha), and insulin stimulation of quiescent cells, to cite only a few treatments, result in phosphorylation of rp S6 and an increase in protein synthesis caused by an increase in the rate of initiation (Thomas et al., 1982). When cells are treated with serum in the presence of cycloheximide, protein synthesis is inhibited but rp S6 phosphorylation is not affected (Thomas et al., 1980). However, when cells are stimulated by serum in the presence of the cAMP phosphodiesterase inhibitor, methylxanthine SQ 20006, both the activation of protein synthesis and rp S6 phosphorylation are blocked. Thomas et al. (1980) concluded that the phosphorylation of rp S6 is required for protein synthesis and therefore, cell proliferation. Individually, EGF and PGF2 alpha do not stimulate rp S6 phosphorylation and protein synthesis to the same extent as serum, but together they act synergistically.

Insulin acts in a fashion similar to that of EGF. The insulin receptor has tyrosine kinase activity which is stimulated when insulin is bound. Microinjection of this receptor kinase into Xenopus oocytes resulted in increased phosphorylation of rp S6 due either to stimulation of a protein-serine kinase or inhibition of a serine phosphatase (Maller et al., 1986).

Cells of the immune system can also be stimulated by certain growth factors which increase rp S6 phosphorylation. In an interleukin-2 (IL-2) dependent lymphocyte clone, IL-2 induced phosphorylation of rp S6 was correlated with an increase in protein synthesis (Evans and Farrar, 1987). Interestingly, the "Tac" receptor upon which IL-2 acts is devoid of tyrosine kinase activity, but there is evidence that the binding of IL-2 to its receptor leads to the activation of protein kinase C (Farrar and Anderson, 1985). Therefore, the IL-2 receptor must mediate rp S6 phosphorylation through a kinase distinct from those activated by receptors with tyrosine kinase activity.

Cells transformed by some viruses also experience an increase in rp S6 phosphorylation as detected by autoradiography of whole protein and of phosphopeptides (Blenis and Erickson, 1984). Murine fibroblasts infected with any one of Rous sarcoma virus, Fujinami sarcoma virus or PRC II sarcoma virus contain elevated levels of

phosphorylated rp S6, as detected by autoradiography, with 95% as phosphoserine and the rest as phosphothreonine. Unlike these avian sarcoma viruses whose transforming proteins possess tyrosine kinase activity, Harvey murine sarcoma virus, also a retrovirus, employs the protein p21, which has no tyrosine kinase activity, as its transforming gene product. Transformation with p21 also causes an increase in rp S6 phosphorylation. In addition, the DNA transforming viruses SV40 and polyoma virus stimulate rp S6 phosphorylation when they infect permissive cells.

Some tumour promoters affect the level of rp S6 phosphorylation upon interaction with cells. For example, chick embryo fibroblasts and murine fibroblasts treated with agents such as phorbol 12-myristate 13-acetate (PMA) exhibit the same high levels of phosphorylated rp S6 that are similar to those produced by growth factors and transforming viruses (Blenis et al., 1984). These authors suggest these effects are probably mediated by protein kinase C.

All of these growth factors, viruses and chemicals have as a common effect the phosphorylation of rp S6. This may occur as the end point of a series of kinases with the initiating events having diverse origins. While the trigger may differ, the phosphorylation of rp S6 seems to be an event shared by all these kinase pathways.

The relevance of this event to the morphological and metabolic changes which accompany transformation has, however, yet to be established.

Rabbit reticulocytes contain at least three kinases that phosphorylate rp S6 *in vitro* and *in vivo*. One is a cAMP-dependent protein kinase and two are cAMP-independent kinases called protease activated kinases I and II. *In vitro*, the former kinase adds 2.0 moles of phosphate per mole of 40S subunit while the latter two add 2.5 moles (Perisic and Traugh, 1983). Other kinases that phosphorylate rp S6 have been isolated from rat brain (LePeuch et al., 1983), and from cells stimulated by EGF (Novak-Hofer and Thomas, 1984), by nerve growth factor (Matsuda and Guroff, 1987) or by fibroblast growth factor (FGF) (Pelech et al., 1986). The latter kinase also phosphorylates the serine residues of a synthetic hepta-peptide that contains the major phosphorylation sites for rat cAMP dependent protein kinase.

The phosphorylation state of ribosomes changes depending on whether the ribosomes are present as subunits, 80S ribosomes or as polyribosomes. In one analysis, 40S subunits and 8.5 ribosomes were found to be phosphorylated to the same extent, but polyribosomes were phosphorylated to a much higher level as demonstrated by autoradiographs of two dimensional gels containing isolated ribosomal proteins (Thomas et al., 1982). The

authors suggested that highly phosphorylated ribosomes have a selective advantage in entering polyribosomes. Burkhard and Traugh (1983) have demonstrated that the *in vitro* phosphorylation of rabbit reticulocyte rp S6 results in selective translation of some synthetic messages. Selective translation of naturally occurring mRNA after the phosphorylation of rp S6 *in vivo* has yet to be demonstrated.

The process of rp S6 phosphorylation and the protein's topographical location on the ribosome have been the subjects of study by several groups. Some of their results provide circumstantial evidence in support of a functional role for rp S6 phosphorylation. One interesting observation is that serum stimulation causes certain tryptic peptides of rp S6 to become phosphorylated in a specific order (Martin-Perez and Thomas, 1983; Martin-Perez et al., 1986). This does not seem compatible with the notion that rp S6 is phosphorylated non-specifically by wandering kinases as implied by Johnson and Warner (1987) (see below). Experiments in which the phosphorylated and dephosphorylated states of rat liver ribosomes were compared showed that a conformational change occurs in the 40S subunit as phosphate groups are added (Kisilevsky et al., 1984). The accessibility of some ribosomal proteins to reductive methylation was altered after

phosphorylation such that rp S3, S4, S7 and S23/24 consistently showed changes in the degree of labelling that paralleled changes in phosphorylation. These ribosomal proteins had previously been shown by cross-linking experiments with 2-iminothiolane to be situated near rp S6. Moreover, rp S6 and rp S3 can be cross-linked to eukaryotic initiation factor 2 (Tolan and Traut, 1981). These observations suggest that the degree of phosphorylation of rp S6 is determined by the growth requirements of the cell and that this phosphorylation plays a role in translation.

Even though several groups believe that rp S6 phosphorylation is part of an important control mechanism in protein synthesis, others do not share such speculation. The equivalent of mammalian rp S6 in yeast is rp S10. This ribosomal protein is believed to be phosphorylated on two adjacent serine residues close to its C terminus. Kruse et al. (1985) and Johnson and Warner (1987) have employed site specific mutagenesis to change both serine residues to alanine in one or both rp S10 coding genes. In a strain mutated in both of the rp S10 coding genes there was no detectable rp S10 phosphorylation and more significantly no observable phenotypic effect. This negative result raises important questions regarding the role of phosphorylation of yeast rp S10 and by implication, of rp S6 in higher organisms.

It serves to highlight the fact that phosphorylation of rp S6 in cells of higher eukaryotes has always been correlated with some other effects. The decisive experiment to demonstrate an essential regulatory function for phosphorylated forms of rp S6 remains to be reported.

e. Organization of genes encoding rRNA and ribosomal proteins

Ribosomal RNA is transcribed from tandemly repeating units of rDNA (Wellauer and Dawid, 1973). In humans, there are approximately 160 of these repeating units per haploid genome (Schmickel, 1973) located on five chromosomes at positions 13p12, 14p12, 15p12, 21p12 and 22p12 (O'Brien, 1987). Each of the units, in the 5' to 3' direction, consists of an external transcribed spacer (ETS)-18S RNA-an internal transcribed spacer (ITS)-5.8S RNA-ITS-28SRNA and a "non-transcribed spacer". The primary transcript is 45S RNA which is processed to yield 5.8S, 18S and 28S RNA (Erikson et al., 1981). The genes coding for 5S RNA are found on chromosome 1 (Steffensen et al., 1974) at positions 1q42 to 1q43 (O'Brien, 1987).

In contrast to the organization of the rRNA genes, those sequences encoding ribosomal proteins do not appear to be located in specific regions of the genome. Southern

blot analysis of mouse genomic DNA, digested with various restriction enzymes and hybridized with labelled mouse ribosomal protein cDNAs, shows that sequences homologous to these probes are present in 7 to 20 copies (Monk et al., 1981). In another experiment, five mouse cDNAs were used as probes in a Southern blot analysis of DNA from a panel of mouse-hamster hybrids containing various portions of mouse chromosomes (D'Eustachio et al., 1981). The results showed that the sequences homologous to the probes were spread over several chromosomes with no obligatory clustering of mouse ribosomal protein gene sequences and rDNA genes.

Recently, the mouse rp gene families representing rp L7 (Klein and Meyuhas, 1984), rp L30 (Weidemann and Perry, 1984), rp L32 (Dudov and Perry, 1984) and rp S16 (Wagner and Perry, 1985) have undergone intensive study. In each case the family is composed of one gene containing multiple introns and several partially or fully processed pseudogenes. Northern blots of polyadenylated RNA probed with intron sequences demonstrate the presence of unprocessed RNA and add support to the hypothesis that it is only the intron-containing genes that are expressed (Wagner and Perry, 1985). Processed pseudogenes may have arisen by reverse transcription of mature ribosomal protein mRNA, a theory that is supporte by the presence of a polyadenylate

tract at the 3' end of one of the rp L32 pseudogenes (Dudov and Perry, 1984). Further analysis of ribosomal protein pseudogene sequence divergence has even led to a proposed order of evolutionary emergence or integration of each of the pseudogenes (Peled-Yalif et al., 1984).

The arrangement of DNA sequences encoding ribosomal proteins in humans has not been studied as extensively as in rat and mouse. Initial studies involving hybridizing 10 rodent and human ribosomal protein encoding cDNAs to human chromosomes revealed that the probes anneal to chromosomes 5, 8 or 17 (Nakamichi et al., 1986). While the authors suggest that ribosomal protein gene sequences may be limited to these chromosomes, it is clear that an insufficient number of probes was used to generalize so broadly.

Comparison of the characterized murine ribosomal protein genes has permitted some generalizations about their structure. Murine ribosomal protein encoding genes are characterized by pyrimidine palindromes at the cap site, by GC-rich regions around the cap site, by the absence of a TATA box in the -20 to -30 region and by relatively short 5' and 3' untranslated regions (Wagner and Perry, 1985; Weidemann and Perry, 1984; Dudov and Perry, 1984). To date, the only human ribosomal protein gene that has been characterized codes for rp S14 (Rhoads et al., 1986). Although it has a TATA sequence at nt -33

to -30, the experiments that are necessary to show it to be crucial to a functional RNA polymerase II promoter have not been published. Like the murine genes, the cap site is surrounded by GC-rich regions, but unlike the other characterized ribosomal protein genes, rp S14 contains Alu family sequences in the third and fourth introns. It is too early to decide if these repetitve sequences may be a general feature of human ribosomal protein genes.

While much is now known of the cDNA sequences of ribosomal proteins of higher eukaryotes there is little data available for the functional genes and some of the generalizations currently proposed may later prove to be erroneous.

f. Ribosomal Assembly

In higher eukaryotes the production of ribosomes occurs mainly in the nucleolus and is completed in the cytoplasm. Nucleoli arise from chromosomal sites called nucleolar organizing regions (NOR) which contain the genes for ribosomal RNA (rRNA). Transcription of rDNA, 5S RNA and rp genes represents the first step in ribosomal biogenesis and involves all three classes of eukaryotic DNA dependent RNA polymerases: RNA polymerase I (pol I) is responsible for the transcription of 45S rDNA, RNA

polymerase II (pol II) for protein coding genes including those encoding ribosomal proteins and small nuclear RNAs (sn RNA) and RNA polymerase III (pol III) for 5S RNA, tRNA genes and for repetitive elements such as Alu family sequences. These enzymes function under the guidance of transcription factors (TF) which may themselves be the subjects of other control systems. Although not completely elucidated, the mechanisms regulating the three RNA polymerases are gradually being uncovered and should contribute to the understanding of cellular growth control. Briefly, rRNA genes of the NOR are found in the fibrillar centre where transcription starts (Sommerville, 1985). Proteins are added to these transcripts in the surrounding dense fibrillar component to form preribosomes which proceed to the granular component where more processing of rRNA occurs. During ribosomal maturation some accessory proteins are involved that do not form part of the final product.

The first step in ribosome formation is rDNA transcription. Immunofluorescence studies have localized pol I to the fibrillar centres of the nucleolus where rRNA transcripts are first seen (Scheer et al., 1983; Scheer and Rose, 1984). The enzyme binds to a promoter region of rDNA that spans -40 to +53 nt of the gene (Muramatsu et al., 1986). To form a functional and accurate initiation complex pol I requires the presence

of at least two transcription factors (TF) known as TFIA and TFID (Mishima et al., 1982). TFID binds to the promoter and is itself subsequently bound by TFIA such that only then may pol I join the initiation complex and start transcription. The TF remain with the promoter as pol I reinitiates several rounds of transcription. Species specificity is conferred by TFID in that it cannot be exchanged between mouse and human (Learned et al., 1985; Grummt et al., 1982; Miesfield and Arnheim, 1984).

As the rDNA genes are transcribed by pol I, pol III begins transcription of the 5S RNA genes. In Xenopus laevis a key part of the 5S promoter is a 34 nt segment lying within the gene. This internal element controls initiation of transcription at a point 50 nt upstream (Sakonju et al., 1980). This architecture is common for class III genes such as tRNA, 7S K and L genes, some viral associated genes and Alu family sequences (Cilberto et al., 1983). In addition to pol III, class III genes also require TFIIIB and TFIIIC and in particular 5S RNA genes need TFIIIA (Segall et al., 1980). TFIIIA binds directly to the non coding strand and in a manner similar to the pol I initiation complex it does not dissociate during several rounds of transcription (Sakonju and Brown, 1982).

Transcription of protein coding genes is initiated by RNA polymerase II in the presence of TFIIA to TFIIE and the first two nucleotides dictated by the template (Dynan and Tjian, 1983; Samuels et al., 1984). This process can be divided into two or three steps (Fire et al., 1984; Hawley and Roeder, 1985) which lead to the formation of an elongation complex. Transcription proceeds while the TF remain behind to prepare for another round of initiation (Workman and Roeder, 1987). Since TFIIID is a TATA box binding factor it will be interesting to determine its role, if any, in transcription of ribosomal protein coding genes as most ribosomal protein genes in higher eukaryotes lack a TATA box (Dudov and Perry, 1984; Wiedemann and Perry, 1984; Wagner and Perry, 1985).

The three transcription processes occur simultaneously as demonstrated by the association of ribosomal, and other proteins, with rRNA before the transcript has been finished (Chooi and Lieoy, 1981). These preribosomal particles continue to be processed in mammalian cell nuclei into three detectable ribosome precusors designated by their sedimentation coefficients as 80S, 55S and 40S (Warner and Soeiro, 1967). The 80S particle contains 45S rRNA and some protein and proceeds along the maturation pathway before cleavage into 55S and 40S particles which contain 32S RNA and 21S precusor

rRNAs, respectively. 5S RNA can be found in the 55S particle making it the precusor of the large subunit while the small subunit is derived from the 40S preribosome (Todorov et al., 1983). Most ribosomal proteins associate with the preribosomes during processing in the nucleolus while a few others are added in the cytoplasm (Lastick, 1980; Todorov et al., 1983).

The predominant proteins in the nucleolus are not ribosomal but are phosphoproteins that may function as helpers or regulators in preribosome assembly. One of these is C23 a 110 kd protein that binds DNA in the nontranscribed spacer region just upstream of the 45S rDNA transcription initiation site (Olson, 1983). Another nucleolar phosphoprotein is B23, a 37 kd protein which is more generally distributed in the RNA containing portions of the nucleolus (Spector et al., 1984).

In addition to the phosphoproteins and the 55S and 40S particles, nucleoli of Xenopus also contain a 65S particle that may be a product of the 55s preribosome (Hugle et al., 1985). Within this 65S preribosome is a 40 kd protein called ribocharin which may act as a transport molecule, assisting in moving the preribosome from the nucleolus to the cytoplasm without ever leaving the nucleus itself. Another abundant protein in the Xenopus nucleolus is a 180 kd acidic protein that is found in transcriptionally active areas (Schmidt-Zachman et al.,

1984). All of these RNA molecules, ribosomal proteins and associated nucleolar proteins are involved in the biogenesis of ribosomes in higher eukaryotes. Control of the synthesis of these elements at the levels of both transcription and translation is under intensive study by several groups.

g. Transcriptional control of ribosomal protein synthesis

In yeast, transcriptional control of rp synthesis has been investigated by the use of fusion genes. One such experiment employed a construction that ligated the 528 nt upstream of the transcription start of rp55-1 to the 3' end of the E. coli lacZ gene (Donovan and Pearson, 1986). The relative rate of synthesis of the fusion protein was compared to that of five ribosomal proteins during a nutritional upshift and found to follow the same kinetics implying that the 528 nt was sufficient to mediate the change in the rate of synthesis. To further define the upstream signals needed for transcription, a fusion gene including 1030 nt upstream of RP 39A and the large 3' fragment of the E. coli lacZ gene was constructed and then subjected to Bal 31 digestion or Xho I linker insertion to generate a series of 5' mutants (Rotenberg and Woolford, 1986). Measurement of beta-galactosidase activity in the mutant clones

showed that three stretches of 5' flanking DNA were necessary for efficient expression of the fusion gene. A computer search of 19 other yeast ribosomal protein genes found that within the 500 nt 5' to their start codons all contained at least one of these consensus sequences and most had all three. The gene encoding elongation factor EF1-alpha also contains one of these sequences in its 5' flank suggesting that transcription of yeast ribosomal protein genes and associated translation factors may be under a global control that can quickly react to changing conditions in the cell (Huet et al., 1985).

Experiments with deletion mutants have also been employed to delineate the transcription control sequences of mouse ribosomal protein genes. A series of plasmids carrying the rp L32 gene and various deletions of the DNA 5' to the transcriptional start site was constructed and transfected separately into primate cells. The transient expression of the altered rp L32 DNA was measured using a transfected but unaltered mouse rp S16 gene as an internal control (Dudov and Perry, 1986). Initially, the authors determined that only the 36 nt 5' to the cap site in the gene were necessary for expression. Later studies with different constructions revealed that the vector used in the first experiments contained an enhancer and that in fact, 79 nt 5' to the gene and the first 31 nt of

the first intron are needed for accurate transcription (R. P. Perry pers. commun. to G. A. Mackie).

The evidence gathered thus far suggests that, in eukaryotes, ribosomal protein gene transcription is regulated by relatively short DNA sequences situated immediately 5' to the genes and in the first intron. This work must be extended to determine if this is a common motif for other mammalian ribosomal protein coding genes.

Post-transcriptional regulation also appears to be an important means of control of ribosomal protein synthesis. In E. coli ribosomal protein synthesis is regulated autogenously. When some ribosomal proteins are produced in excess, these ribosomal proteins bind to their own messages and inhibit their translation (Nomura et al., 1984; Lindahl and Zengel, 1986). In eukaryotes the presence of a nuclear membrane, introns and splicesomes allows for more opportunities to establish points of control for ribosomal protein synthesis.

Yeast exhibits at least three types of post-transcriptional regulation of rp synthesis. One type was demonstrated when a high copy number plasmid carrying the gene encoding rp S10 was transfected into yeast cells (Warner et al., 1985). An increase in the level of rp S10 mRNA was accompanied by a corresponding increase in the level of rp S10 protein in a short pulse. After 30 min the message was still produced at high levels, but

after a chase the amount of rp S10 returned to normal implying an increase in protein turnover compensating for the excess message. A similar experiment in which yeast was transformed with a plasmid carrying the gene encoding rp L29 resulted in an accumulation of both the rp L29 precursor RNA and the mature mRNA while the protein was present in stoichiometric amounts. Excess rp L29 mRNA was found in the polyribosomal fraction implying that there were fewer ribosomes per message. in other words, there was modulation of translational efficiency of the "excess" L29 mRNA (Warner et al., 1985). Another form of post-transcriptional regulation was identified when the gene encoding yeast rp L32 was intoduced into a recipient strain on a high copy number plasmid. Although rp L32 precusor RNA accumulated, the message did not, leading the authors to conclude that the protein could bind to its own primary transcript and inhibit splicing (Dabeva et al., 1986). Whether or not these are the only mechanisms of post-transcriptional regulation awaits further investigation; it is interesting, however, to note that yeast does not employ just one method for all ribosomal proteins.

Xenopus rp L1 may be controlled like yeast rp L32 as its primary transcript accumulates when the gene for L1 is microinjected into oocytes (Bozzoni et al., 1984). More recent experiments have demonstrated that intron 3

in the L1 gene is a target of regulation. A block to splicing of this intron allows endonucleases to attack specific regions of the precusor RNA and destabilize it (Caffarelli et al., 1987).

Post-transcriptional regulation of ribosomal protein synthesis has been demonstrated in multicellular eukaryotes as well as yeast. Pierandrei-Amaldi et al. (1985, have discovered that the polyribosome/mRNP distribution of ribosomal protein mRNA in Xenopus laevis may be involved in determining how much rp mRNA is available for translation. These investigators propose that autogenous regulation may permit the mRNA to be mobilized quickly if the cell requires new ribosomes. Post-transcriptional regulation of ribosomal protein synthesis has also been demonstrated in mouse P1798 lymphosarcoma cells which were treated with glucocorticoids to suppress rRNA synthesis. In these cells, despite an accumulation of rp mRNA, the proportion of ribosomal protein mRNA associated with polyribosomes ranged from 22 to 38% as compared with 59 to 76% in untreated cells. Non-ribosomal protein mRNAs remained essentially unaffected by the hormone treatment. These results suggest that glucocorticoid treatment causes a selective decrease in the association of ribosomal protein mRNA with polysomes (Meyuhas et al., 1987). The authors state that translational efficiency is greatly

increased in an mRNA that is already being translated by another ribosome. They speculate that this could be due to a change in conformation of the message or it may be that glucocorticoids stablize mRNP preventing their mobilization into polysomes. Whichever is the case, 5 of 6 ribosomal protein mRNAs examined followed the same pattern after glucocorticoid treatment suggesting a coordinate translational control of ribosomal protein synthesis.

In a set of stable transformants of CHO cells transfected with a plasmid carrying the human gene encoding rp S14 there were up to 35 fold increases in the relative amount of rp S14 mRNA but only slight variations in the amount of S14 protein (Rhoads and Roufa, 1987). The excess message, found in polyribosomes, could not have been translated efficiently or "excess" S14 was degraded rapidly or both.

Although much has yet to be learned about the regulation of ribosomal biogenesis, it is now known that there are several steps along the pathway that can serve as control points. Transcriptional and post-transcriptional regulation of ribosome synthesis may eventually be found to be related in a complex process responding to changing metabolic states in individual cells.

h. Cloning of rp cDNAs and genes

Ribosomal proteins are usually small basic proteins encoded by relatively small mRNA molecules (Warner, 1979). As discussed earlier, the ribosomal protein messages isolated to date from higher eukaryotes usually have 5' untranslated regions of 70 to 90 nt and short 3' untranslated regions of 30 to 50 nt. Taken together these data confirm that most ribosomal protein mRNAs sediment at less than 12S in sucrose gradients and are therefore separable from the bulk of eukaryotic messages at about 18S to 20S. This feature has been exploited by several groups who have isolated sucrose gradient fractions enriched for ribosomal protein mRNA and used them to identify ribosomal protein cDNAs. Short 5'or 3' untranslated regions may reflect the choice of small mRNAs that this method selects and may not be a characteristic of all ribosomal protein messages.

Fractions of lower molecular weight yeast mRNA have been translated in cell-free systems and the products separated by two-dimensional gel electrophoresis to find those containing mRNAs coding for ribosomal proteins (Woolford et al., 1979). These fractions were radiolabelled and used to probe a yeast genomic library. Putative ribosomal protein genes were identified by

hybrid-selected translation to determine which of the genomic clones contained complete protein encoding genes.

A further refinement of this approach was developed to isolate mouse ribosomal protein coding cDNA clones. The enriched RNA fractions were used to construct a cDNA library which was screened by hybrid selection and translation (Meyuhas and Perry, 1980). The proteins from in vitro translation were electrophoresed on two dimensional gels and their mobilities compared to marker ribosomal proteins. This protocol has been used successfully in several model organisms including rat (Kuwano et al., 1985) and Xenopus laevis (Bozzoni et al., 1981).

Another approach relies on the availability of cDNA or genomic sequences encoding ribosomal proteins in one species to be used as probes to screen the libraries, cDNA or genomic, of another species. As discussed above, codon usage and the degree of amino acid divergence between species are important constraints on the success of this method. Thus, sequences from more closely related species are more likely to cross hybridize than more distantly related species. For example, portions of the cDNA clone encoding CHO rp S14 have been used to screen a human cDNA library to identify the corresponding human clone (Rhoads et al., 1986). Similarly, a clone encoding rat rp S11 has been used to isolate a human rp S11 cDNA

clone from a human fibroblast library (Lott and Mackie, 1988). The isolation of DNA sequences in Chinese Hamster Ovary cells and in human genomic DNA that encode RNA polymerase II provides an example of cross-hybridization in more distantly related species. These sequences were identified with a 4.2 kb probe of Drosophila melanogaster DNA that contained part of the coding sequence of RNA polymerase II (Ingles et al., 1983).

The isolation of protein coding genes from recombinant libraries can also be achieved by using antibodies to detect antigen produced by particular clones. Both plasmid (Helfman et al., 1983) and phage (Young and Davis, 1983) expression vectors have been used for this purpose. This approach has been used successfully to isolate a clone from a lambda gt11 expression library that encodes human ribosomal protein S6 (Heinze et al., 1988; published while this thesis was in preparation).

The amino and carboxy terminal sequences of many higher eukaryotic ribosomal proteins have been elucidated by amino acid sequence analysis. These data can be used to design heterologous oligonucleotide probes that can account for degeneracy in the genetic code. To increase a probe's specificity the least degenerate stretch of amino acid sequence should be considered. This strategy has been used to isolate ribosomal protein coding cDNAs from

rat (Chan et al., 1987) and human (Lott and Mackie, 1988b) libraries. Therefore, the investigator has at his disposal many methods by which ribosomal protein encoding cDNA clones can be isolated. While some of these are predisposed to the isolation of low MW rp coding cDNA, others can be used to isolate specific clones of interest.

Finally, another source of mammalian ribosomal protein encoding cDNA is accidental discovery. Cloned cDNAs for ribosomal proteins have been discovered in searches for other cDNAs after computer searches of gene banks with newly determined DNA sequences. Ribosomal protein cDNAs found in this manner have included mouse S6 (Lalanne et al., 1987), the human homologue of yeast L44 (Davies et al., 1986 and the human homologue of E. coli L3 (Ou et al., 1987).

1. Objectives

In 1982 virtually no data were available on ribosomal protein genes in higher eukaryotes and very little was known about cDNA clones. Even in 1988 fewer than one quarter of all cytoplasmic ribosomal protein cDNAs have been cloned from all higher eukaryotes taken together. Coordination and growth control of ribosomal

proteins is poorly understood in yeast and not at all in higher organisms.

This project consisted of several short range and long range goals. The immediate objective was to obtain one or more human cDNA or genomic clones that encode ribosomal proteins. Along with structural data these clones could supply evolutionary information. Comparisons of DNA sequences encoding ribosomal proteins from different organisms may reveal some of the effects of evolutionary pressures on both the DNA coding sequence and on the predicted amino acid sequence. In addition, the untranslated regions residing 5' and 3' to the coding region may harbour clues to post-transcriptional regulation. To be maximally useful, such comparisons require access to full-length coding sequences.

Isolation of human cloned cDNAs or genomic clones would facilitate the long range goals of the project. Briefly, we had planned to employ one or more clones in transfection experiments to achieve an imbalanced gene dosage in recipient cells. Study of precursor RNA, mRNA and protein levels would reveal whether or not cells impose coordinate control at the level of translation or employ other regulatory mechanisms. Cloned cDNA inserts could be ligated into a vector that supplies a promoter and exogenous splice junction (Okayama and Berg, 1983). In contrast, ribosomal protein genes that contain all the

necessary control elements would not require the exogenous signals of an expression vector. Only recently have some of the regulatory features of a small number of mammalian ribosomal protein genes been recognized (R. Perry pers. commun. to G. A. Mackie; 1988). For example, transcription signals for murine rp L30 and L32 are located in the 70 bp 5' to the gene while faithful processing of the primary transcript requires the presence of the first intron.

Ribosomal protein S6 was chosen as the principal target of this investigation because of its implied role in growth regulation. Although rp S6 is one of the more interesting ribosomal proteins its cloned cDNA has been one of the more challenging to isolate. Thus considerable time and effort was invested in different approaches to isolating the full-length coding sequence. The role of mammalian rp S6 phosphorylation in growth regulation can be studied in the future by employing techniques of modern mutagenesis on the cloned cDNA. Altering appropriate codons in the distal portion of the coding region would allow production of mutant rp S6 that could not be phosphorylated to the same degree as the native protein. The phenotypic effects of such changes, if any, may reveal the function of rp S6.

Cloned cDNAs encoding human rp S11 were also isolated as a check on our techniques. They have provided

useful information on the control of stability of rp mRNAs (P. C. Wong, pers. commun.).

CHAPTER 2

ATTEMPTS TO ISOLATE HUMAN RIBOSOMAL PROTEIN CODING SEQUENCES BY CROSS HYBRIDIZATION

2.1 Introduction

Several cDNA and a few genomic clones that encode ribosomal proteins in higher eukaryotes have now been isolated as discussed in Chapter 1. In most cases the strategy was based on the observation that one can enrich mRNA encoding low molecular weight ribosomal proteins by fractionating total mRNA by sucrose gradient centrifugation. Suitable fractions were used to generate cDNA libraries that were screened by various methods to identify cloned ribosomal protein cDNAs (Meyuhas and Perry, 1980). By design this method should lead only to the isolation of cDNAs encoding low molecular weight ribosomal proteins. Ribosomal proteins whose molecular weights are 28 kd or more will be encoded by mRNAs in the range of 900 nt or greater putting them in gradient fractions which also contain many other cellular mRNAs. This limits the degree to which the total mRNA can be enriched for ribosomal protein coding mRNAs. We were interested in cloning human ribosomal protein S6 whose size was expected to be 30 kd, corresponding to approximately 270 aa residues. We expected, therefore, that its mRNA would be approximately 1000 nt depending on the length of the polyadenylated tail. Since a human cDNA representing rp S6 would be virtually impossible to locate by hybrid-selected translation without prior

enrichment and since there were no ribosomal protein S6 cDNA clones available from other higher eukaryotes, the probe selected was obtained from one of the yeast genes encoding rp S10. Yeast ribosomal protein S10 is apparently the functional equivalent of mammalian ribosomal protein S6 (Zinker and Warner, 1976).

Others have isolated particular cDNAs and/or genes by cross-hybridization techniques even though the probe and target species may be separated by a large distance on an evolutionary scale. Ingles et al. (1983), were able to identify sequences encoding an RNA polymerase II subunit in Chinese Hamster Ovary cell DNA, Hela cell DNA and in human placental DNA with a probe consisting of a 32P-labeled DNA fragment from a gene encoding a RNA polymerase II subunit in Drosophila melanogaster. Cross-hybridization has also been employed to isolate a mouse DNA fragment which possesses a high degree of homology to the Drosophila ribosomal protein 49 gene (Mulligan and Hackett, 1984).

Another successful approach to cross hybridization employed 32P-labeled RNA transcribed from a bovine rhodopsin cDNA clone to identify homologous transcripts in a Northern transfer of Drosophila melanogaster RNA (Zuker et al., 1985). An interesting feature of this work was the inability of 32P nick-translated bovine rhodopsin cDNA to detect homologous sequences in Drosophlia DNA or

RNA. RNA-RNA hybrids are more stable than RNA-DNA and DNA-DNA duplexes (Casey and Davidson, 1977) suggesting that the degree of DNA sequence homology of the bovine and Drosophila rhodopsin coding regions is much less than 100% (in fact about 30%) and lies between that detected by Bovine DNA-Drosophila RNA hybrids and Bovine RNA-Drosophila RNA hybrids.

In an effort to isolate a clone encoding human ribosomal protein S6, duplicate lifts of a human genomic library were screened simultaneously with a yeast ribosomal protein S10 probe and with a probe from fractionated rat liver polyadenylated RNA of approximately 1000 nt. Clones apparently positive for both probes were isolated for further study.

In additional experiments, a human cDNA library was screened with a cDNA clone encoding rat rp S11 (Tanaka et al., 1985) so that the human homolog of S11 could be isolated as a cDNA both to test our techniques and for use in regulatory experiments (P. Wong and G. A. Mackie, in preparation).

2.2 Materials and Methods

a. Human cDNA and Genomic Libraries

Human cDNA and genomic libraries were obtained from other laboratories as detailed below.

A human genomic library derived from fetal liver and carried by lambda Charon 4A phage (Lawn et al., 1978) was obtained from D. J. Fujita. The host bacterium was E. coli DP50supF and was grown on NZYM-DT (10 gm NZ amine, 5 gm NaCl, 5 gm yeast extract, 2 gm magnesium sulfate, 0.1 gm diaminopimelic acid and 0.05 gm thymidine per litre) for growth.

An adult human liver cDNA library was purchased from Clontech Laboratories, Inc. (Palo Alto, California). The product profile stated that the library consisted of 150,000 independent clones carried in lambda gt10. This phage was grown in E. coli C600 hfl rk-, mk+ as host on LB medium (10 gm tryptone, 5 gm yeast extract and 10 gm NaCl per litre).

Another cDNA library was obtained from Dr. P. Berg. The mRNA source for construction of the library was an SV40-transformed human fibroblast cell line GM 637 and the corresponding cDNAs were inserted into the expression vector pCD (Okayama and Berg, 1983). The cellular host

for this library was E. coli X1776 which grows in X-medium (25 gm tryptone, 7.5 gm of yeast extract and 20 ml of 1 M Tris.HCl(pH 7.5), 5 ml of 1 M magnesium chloride, 10 ml of 1% diaminopimelic acid, 10 ml of 0.4% thymidine and 25 ml of 20% glucose per litre).

b. Isolation of Rat Liver and Human Placental RNA

Rat livers were obtained from freshly sacrificed rats maintained on normal diets. Human placentas were obtained from deliveries by Caesarian section and were taken as quickly as possible on ice from the delivery room to the laboratory. The method of RNA isolation is based on two procedures described by Maniatis et al., (1982). Usually, 50 gm of tissue were homogenized in 250 ml of 4 M guanidinium isothiocyanate and then extracted with 250 to 300 ml of phenol at 60 C. The aqueous phase was recovered and then extracted once with phenol/chloroform/isoamyl alcohol (25:24:1) and once with chloroform/isoamyl alcohol (24:1). The RNA was precipitated by adding 2.5 volumes of ethanol and storing at -20 C for at least 1 to 2 h. RNA was recovered by centrifugation and then dissolved in 60 ml of TE (10 mM Tris.HCl and 1 mM EDTA pH 7.4). 1 gm of cesium chloride was added to each 2.5 ml of the RNA solution. 10 ml of this solution was layered onto a 1.2 ml cushion

containing 5.7 M cesium chloride, 100 mM EDTA (pH 7.4) in tubes appropriate for a SW 50.1 rotor. The RNA solution was centrifuged at 35,000 RPM for 12 h at 20 C. The supernatant was discarded and the RNA pellet was dissolved in a solution of 10 mM Tris.HCl pH 7.4), 5 mM EDTA and 1% SDS. To this solution 0.1 volume of 3 M sodium acetate (pH 5.2) was added followed by 2.5 volumes of ethanol. The RNA was stored at -20 C for at least 2 h and then recovered by centrifugation.
Polyadenylated RNA was selected using affinity chromatography on a 1 ml column of oligo (dT) cellulose (Maniatis et al., 1982).

c. Probes for Screening Human Libraries

Two plasmids were obtained that each carried one of the genes coding for rp S10 in Saccharomyces cerevesiae (J. R. Warner pers. commun. to G. A. Mackie; refer to Figure 2.1). The plasmid RPgG consists of a 3.6 kb Hind III to Bam HI fragment of yeast DNA carrying one of the rp S10 genes inserted into pBR322 . The other plasmid, RPgE, contains a 4.5 kb Bam HI to Bam HI fragment of yeast DNA that carries the other rp S10 gene inserted into pBR322. These plasmids are represented in Figure 2.1 along with a 1.9 kb Eco RI to Eco RI fragment of yeast DNA labelled fRPg which contains the 3' end of the rp S10

gene in RPgG. fRPg was used to screen the human genomic library.

The genomic library was also screened with size fractionated polyadenylated rat liver RNA. 100 to 200 ug of polyadenylated RNA that had been selected by oligo (dT) cellulose affinity chromatography (Maniatis et al., 1982) was taken up in 300 ul of 5mM methyl mercuric hydroxide. This sample was layered onto a 12 ml 5% to 20% sucrose gradient that contained 10 mM Tris.HCl (pH 7.4), 0.5% SDS and 1 mM EDTA (Meyuhas and Perry, 1980). Each gradient was centrifuged at 28,000 RPM for 16h in a SW 40 rotor at room temperature. The gradient was fractionated into 500 ul fractions whose absorbance at 260 nm was determined. Appropriate fractions were pooled for probe.

A cloned cDNA inserted at the Pst I site in pBR322 encoding rat rp S11 was obtained from K. Ogata (Taneka et al., 1985) and used for probe.

d. Radiolabelling of DNA Probes

Two procedures were used to label DNA probes with radioactive nucleotides. One of these was nick translation in which about 0.5 ug of DNA was nicked and repaired in the presence of 10 to 20 uCi of [alpha-32P] dNTP using a modification of the procedure of Rigby et al. (1977) and a kit purchased from Bethesda Research

Laboratories. The other procedure was random priming in which DNA fragments, usually from within the coding regions of cDNA clones, were labelled with [alpha-32P] dCTP using the method of Feinberg and Vogelstein (1983).

e. End Labelling of RNA

RNA was partially hydrolyzed by incubation at 90 C for 30 min and then transferred to a microcentrifuge tube containing 50 uCi of [gamma-32P]-ATP. 10 ul of 10X polynucleotide kinase buffer (500 mM Tris(pH 7.6), 100 mM magnesium chloride, 1 mM EDTA (pH 7.6) and 5 mM DTT) and 5 units of polynucleotide kinase were added to the tube making the final volume 100 ul. Incubation proceeded at 37 C for 30 min. The reaction was stopped with 20 mM EDTA (pH 8.0) and extracted with phenol/chloroform/isoamyl alcohol. Labelled RNA was made 2 M in ammonium acetate, was precipitated with 2.5 volumes of ethanol, and was recovered by centrifugation. The RNA was reprecipitated to remove unincorporated nucleotides.

f. Removal of Free Nucleotides After Labelling

The following procedure was used after some of the nick translations, all random priming and all labelling of oligonucleotide probes (see Section 3.2). Sephadex G-

25 superfine was allowed to swell overnight in sterile water. The next day, the Sephadex was poured to a height of 10 cm in a 10 ml disposable pipette and equilibrated with 10 ml of column buffer (20 mM Tris.HCl (pH 8.0), 2 mM EDTA, 100 mM sodium acetate and 50 ug/ml yeast RNA). The labelled probe was dissolved in 100 ul of tube wash (90 ul column buffer, 5 ul glycerol and 5 ul of 0.5 M ammonium sulfate, loaded on the column and eluted with column buffer. Fractions of 8 drops were collected and counted by Cerenkov radiation. The probe fractions were pooled and then precipitated with 2.5 volumes of ethanol to concentrate the nucleic acids.

g. Screening of the Human Genomic Library

The screening procedure was based on the plaque transfer method of Benton and Davis (1977). 10 ml of NZYM-DT were inoculated with 100 ul of an overnight culture of DP50supF and was grown to an OD600 of 1.0. 300 ul of the DP50supF culture were infected with 50,000 PFU from the genomic library and then incubated at 37 C for 20 min to allow the phage to adsorb to the bacteria. Typically, 6 such inocula would be set up so that a screening would involve 300,000 plaques from an aliquot of the library that contained 1 million PFU. 7 ml of molten 0.75% agarose (50 C) were added to each inoculum

and then poured onto 150 mm plates containing 1.5% NZYM-DT agar. These plates were incubated at 37 C for up to 16 h such that they would be just short of confluent lysis. The plates were chilled at 4.0 C for 1 h and then numbered nitrocellulose filters (Schleicher and Schuell, Inc.) were placed on the top agarose for 30 sec. Asymmetric reference marks were made through the filter and into the agar with an 18 gauge needle. Duplicate filters were prepared similarly with the exception that the filters contacted the plate for approximately 60 sec. The nitrocellulose filters were placed on a succession of plates containing moist 3MM paper (Whatman International, Ltd.) as follows; denaturing solution (1.5 M NaCl and 0.5 M NaOH) for 60 sec; neutralizing solution (0.5 M Tris.HCl (pH 8.0) and 1.5 M NaCl) for 5 min; and 2 X SSC for 2 to 5 min. The filters were air dried and then baked in a vacuum oven at 80 C for 2 h.

Filters were prehybridized for 16 h at 37 C in a solution containing 50% formamide, 5 X Denhardt's solution (Denhardt, 1966), 5 X SSC, 0.1% SDS and 100 ug/ml of sonicated herring sperm DNA. The filters were subsequently drained and then 75 ml of fresh hybridization solution were added. Denatured, [alpha-32P] dATP-nick translated fRPg was added at 9 million cpm to one set of filters. The duplicate filters were probed in a similar manner with 6 million cpm of end-labelled RNA

from the pooled sucrose gradient fractions. Hybridization continued for 48 h after which filters were washed twice for 1.5 h in a solution of 1 X SSC and 0.1% SDS at 37 C. The filters were air-dried and and exposed to Kodak XAR film for 48 to 72 h at -70 C (see Figure 2.3). Plaques that appeared to anneal to both probes were picked for further examina on and rescreening. DNA from clones of interest was isolated from plate lysates (Maniatis et al., 1982), digested with Eco RI and analyzed by electrophoresis on agarose gels for characterization and isolation of inserts.

h. Screening of the Human cDNA Library in Lambda gt10

The lambda gt10 human adult liver cDNA library (Clontech, Inc.) was screened in much the same way as the genomic library with some modifications. The phage host was E. coli C600 rk-, mk+, hfl and was grown on LB agar. In one series of experiments the library was probed with a nick translated fragment of human genomic DNA. The probe was derived from the Eco RI to Bam HI insert of pJL-002 (see Figure 2.4) and was used to screen up to 100,000 phage at a time. Prehybridization lasted for 8 to 10 h at 42 C followed by hybridization at the same temperature for 16 h. Filters were washed twice at 50 C

in 1 X SSC and 0.1% SDS for 45 min. After air drying the filters were exposed to Kodak XR film for 20 h.

In separate experiments, the library was screened with a nick-translated yeast probe for rp S10. The probes consisted of an Eco RI to Hind III fragment of RPgE containing 650 nt of the distal portion of the gene and an Eco RI to Eco RI fragment containing 650 nt of the proximal portion of the gene (Figure 2.1). For each plate of 50,000 phage, 4 filters w[illegible] lifted and prepared for hybridization. Prehybridization and hybridization continued at 42 C for 6 h and 63 h respectively in solutions containing 20% formamide, 5 X Denhardt's solution (Denhardt, 1966), 5 X SSC, 50 ug/ml herring sperm DNA, 10 mM PIPES (pH 6.[illegible]) and 0.1% SDS. In each case 5 million cpm of probe were added to give a final concentration of 100,000 cpm/ml. After hybridization the filters were washed four times for 20 min in 0.2 X SSC and 0.1% SDS. Plaques that annealed to both probes on both sets of filters were picked and rescreened. DNA from clones of interest was isolated from plate lysates (Maniatis et al., 1982), digested with Eco RI and analyzed by electrophoresis on agarose gels.

1. Screening of the Plasmid cDNA Library

The human fibroblast cDNA library in pCD (Okayama and Berg, 1983) was grown on X 1776 agar containing 40 ug/ml of ampicillin. Twenty-two 90mm plates each having approximately 1200 colonies were grown and then replicated to nitrocellulose filters. The filters were denatured, neutralized, vacuum dried and then washed overnight at 65 C in 3 X SSC and 0.1% SDS. This served to remove much of the bacterial debris on the filters without significantly affecting the retention of plasmid DNA. The filters were prehybridized in 75 ml of the 50% formamide solution described above at 42 C for 24 h after which 10 million cpm of probe were added and hybridized for 24 h at 42 C. The probe was a 208 bp [alpha-32P] nick translated Bam HI to Hinf I DNA fragment from the insert in pRS-11-2 that codes for rat rp S11 (Tanaka et al., 1985). Filters were washed 4 X in 1X SSC and 0.1% SDS, air dried and then exposed to Kodak XAR film for 24 h (Figure 2.5). Apparent positives were rescreened and then streaked on agar plates for single colonies. Plasmids were extracted from 1.5 ml cultures of individual clones by the alkaline lysis procedure (Maniatis et al., 1982) prior to analysis.

j. Restriction Endonuclease Digestion of DNA

Restriction endonucleases were purchased from New England Biolabs, Bethesda Research Labs, Boehringer-Mannheim or Pharmacia. All digestions with restriction endonucleases were performed in buffers and at temperatures recommended by the supplier at an enzyme to DNA ratio of approximately 2 units to 1 ug of DNA. Digestions in which the fragments of interest were greater than 700 bp were electrophoresed on 0.7% to 1.0% agarose gels in TAE (20 mM Tris-acetate, 40 mM Na acetate and 1 mM EDTA plus acetic acid to pH 8). DNA bands were visualized by staining the gel with 0.5 ug/ml of ethidium bromide and viewing with ultraviolet light. DNA fragments less than 700 bp were electrophoresed on 6% acrylamide gels in TBE (89 mM Tris-borate, 89 mM Boric acid and 2 mM EDTA (pH 8.0)) and visualized by staining with ethidium bromide and viewing under ultraviolet light.

k. Isolation and Subcloning of DNA Fragments

Large DNA fragments were isolated from agarose gels by electroelution. Once the DNA was electrophoresed to the desired distance and the gel stained with ethidium bromide, a 3 mm wide slot was cut in the gel just ahead

of the band of interest. The slot was lined with dialysis tubing and filled with TAE buffer. Electrophoresis at 100 volts continued until the band was entirely in the well or trapped on the tubing. The current was reversed for 1 min and then the contents of the well were removed to a microcentrifuge tube. DNA was extracted sequentially with phenol, phenol/chloroform and chloroform, then made 0.2 M in NaOAc and finally precipitated for at least 2 h at -20 C with 2 volumes of ethanol.

Smaller DNA fragments were eluted from acrylamide gels. The band of interest was excised from the gel and introduced into a 1000 ul pipette tip that had been sealed at the apex with sterile, siliconized glass wool and parafin wax. A sterile glass rod was used to macerate the gel fragment into a paste after which 600 ul of elution buffer (500 mM ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS and 10 ug/ml of E. coli tRNA) was added. The tube was sealed and placed at 37 C overnight to allow the DNA to diffuse into the buffer. After recovering the buffer, the DNA was extracted sequentially with phenol, phenol/chlorofrom and chloroform, made 0.2 M in NaOAc and precipitated with two volumes of ethanol.

Some of the DNA inserts isolated were subcloned into plasmids pBR322, pBR325 and pSP65 using standard methods (Bolivar, 1978; Melton et al., 1984).

Large quantities of plasmids were prepared using a modification of the method of Mandel and Higa (1970).

l. Nucleotide Sequence Analysis

Nucleotide sequencing was carried out using the dideoxy chain termination method after subcloning appropriate restriction fragments into M 13-based cloning vectors (Biggin et al., 1983; Messing, 1983).

m. Analysis of Cellular RNA

Human placental polyadenylated RNA was isolated as described previously in this chapter. RNA was denatured with methyl mercuric hydroxide (Bailey and Davidson, 1976), electrophoresed at 5 ug/lane in a 1.2% agarose gel, transferred to a Biodyne A nylon membrane (Pall, Inc.) and fixed to the support by the method of Southern (1975). Blots were prehybridized for 16 h and then hybridized for 24 h at 42 C with the exception of the RPgE probe which was hybridized at 35 C. The DNA used as probe was nick-translated in the presence of [alpha-32P] dCTP and added to the hybridization solution for a final concentration of 300,000 cpm/ml. The probes were the following: 1) a Hind III cDNA fragment from the coding region of chicken actin obtained from Dr. D. T. Denhardt,

2) a 650 nt Eco RI to Hind III fragment of yeast DNA including the 3' end of the rp S10 coding gene from RPgE. 3) the Eco RI to Bam HI insert from pJL-0.6 (see Results), 4) the Bam HI to Bam HI insert from pJL-1.6 and 5) the Bam HI to Eco RI insert from pJL-4.6. The strips were washed 4 X 15 min in 2 X SSC and 0.1% SDS and then exposed to Kodak XR film for 72 h.

2.3 Results

a. Screening a Human Genomic Library for the Gene(s) for Ribosomal Protein S6

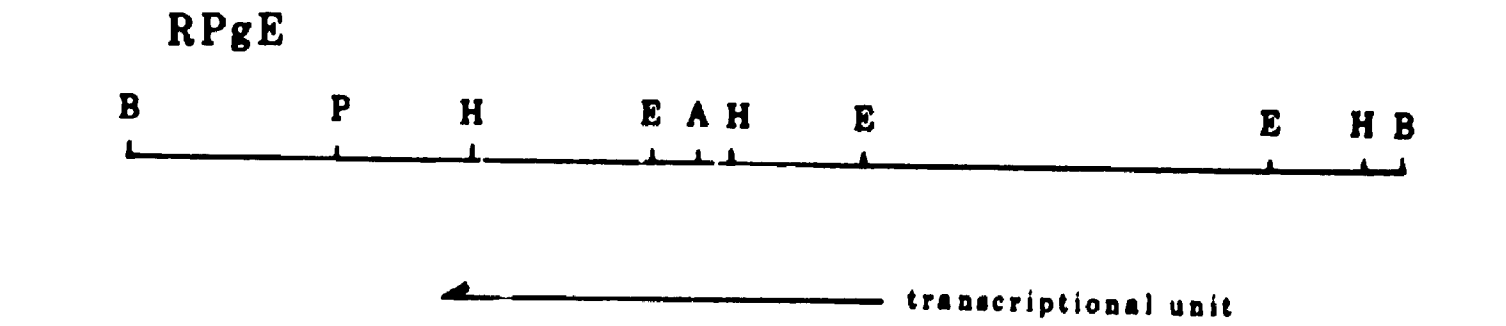
The goal of these experiments was to obtain a gene encoding human rp S6. Since no higher eukaryotic rp S6 probe was available DNA encoding the yeast rp S10 gene was selected to probe various human genomic and cDNA libraries. An existing human genomic library was employed as it had been well-characterized and had already served as the source for the successful isolation of human genes (Lawn et al., 1978). No evidence existed to suggest that the library did not accurately represent the human genome. The human cDNA libraries that proved to be most informative subsequently had previously been screened successfully. Human cloned cDNAs containing complete 3' untranslated regions, full coding sequences and portions of the 5' untranslated regions have been isolated from both a human fibroblast library (Okayama and Berg, 1983; Jolly et al., 1983) and a human placental library (Ullrich et al., 1984).

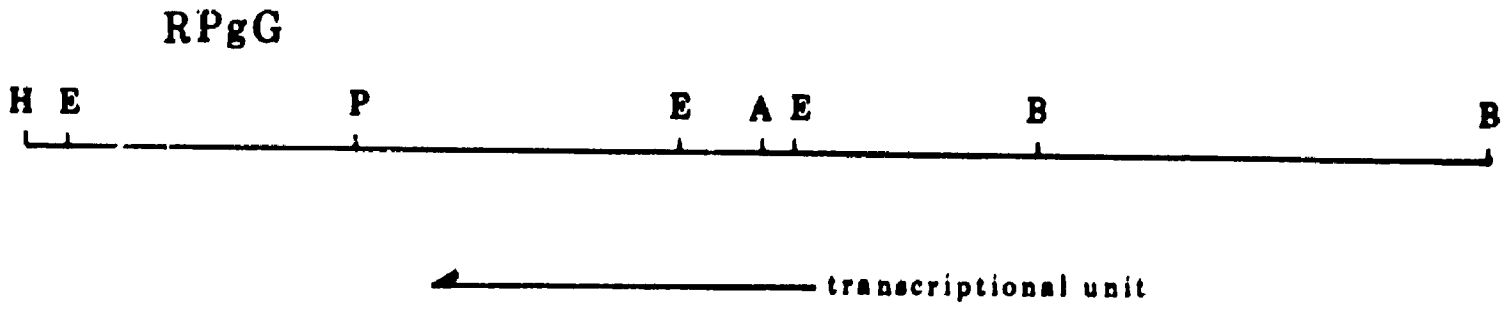
There is no evidence that ribosomal proteins are tissue-specific; indeed, every cDNA library that attempts to represent all cellular mRNAs will necessarily contain

ribosomal protein encoding cloned cDNAs. We decided, therefore, that there would be little benefit in constructing our own libraries when well-characterized libraries were already available. In addition, several experienced colleagues recommended this approach.

A comparison of the published amino acid sequence (predicted by the DNA sequence) of rp S10 from S. carlsbergensis (Leer et al., 1982) and the N terminal sequence of rat rp S6 (Wool, 1979) revealed that there was 60% amino acid homology between the two at the N termini. Due to the degeneracy of the genetic code, the DNA homology could be even less than 60%. In screening a human genomic library with a probe that was probably not homologous to the target, one would expect the probe to occasionally anneal to plaques not carrying any rp S6 coding DNA. One replica of a human genomic library was probed with a nick-translated fragment of yeast DNA: fRPg in Figure 2.1. In an effort to limit the number of false positives, duplicate filters were probed with end labelled rat liver RNA from a sucrose gradient fraction containing RNA sufficiently large to encode ribosomal protein S6 (Figure 2.2). Only those plaques that appeared to be positive to both probes were chosen for further analysis (Figure 2.3).

Figure 2.1. Restriction maps of DNA inserts from RPgE and RPgG. The DNA inserts were isolated from Saccharomyces cerevesiae and each contains one of the rp S10-coding genes (J. R. Warner pers. commun. to G. A. Mackie). fRPg is aligned with RPgG and represents the yeast DNA fragment from RPgG used to probe the human genomic library (Lawn et al., 1978). Arrows indicate the extent of the rp S10 transcriptional units (J. R. Warner pers. commun. to G. A. Mackie; Johnson and Warner, 1987). The restriction sites are: B, Bam HI; P, Pst I; H, Hind III; E, Eco RI; A, Acc I.


RPgE
B
P
H
E
A
H
E
E
H
B
transcriptional unit


RPgG
H
E
P
E
A
E
B
B
transcriptional unit


fRPg

Figure 2.2. Profile of rat liver polyadenylated RNA fractionated on a sucrose gradient. 100 ug of polyadenylated rat liver RNA was fractionated on a 5-20% gradient so that 28S rRNA and part of the 18S rRNA were pelleted. The gradient was fractionated and the absorbance at 260 nm was monitored. Fractions 4, 5 and 6 were pooled and used as probe. Arrows indicate the extent of these fractions on the absorbance profile.

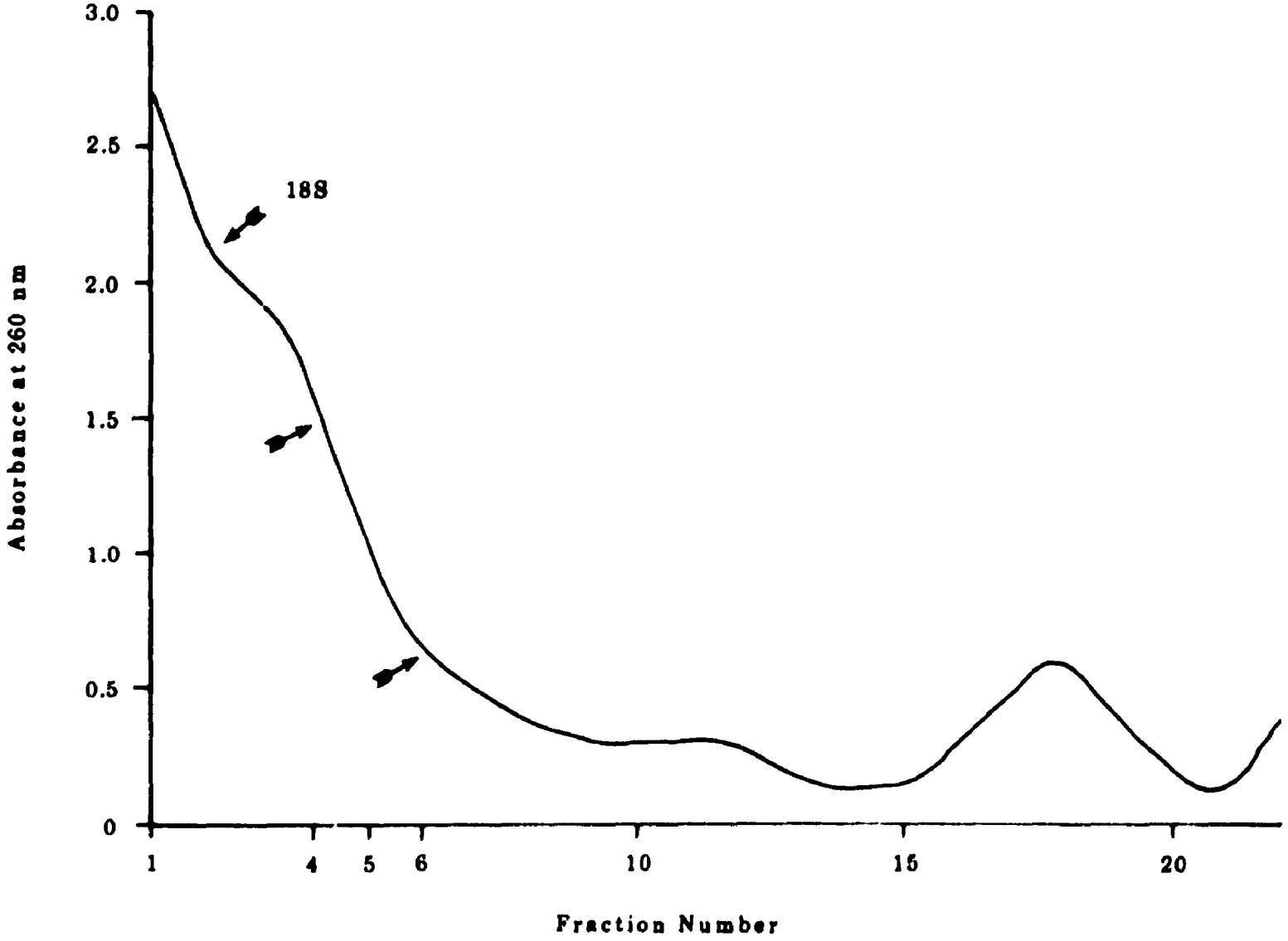

3.0
2.5
2.0
1.5
1.0
0.5
0
Absorbance at 260 nm
18S
1
4
5
6
10
15
20
Fraction Number

Figure 2.3. Autoradiographs of duplicate nitrocellulose filter plaque lifts that were screened with different probes. An aliquot of a human genomic library (Lawn et al., 1978) was screened with a fragment of yeast DNA (panel a) that contained the 3' end of one of the yeast rp S10-coding genes (fRPg; refer to Figure 2.1) and also with a rat liver RNA fraction (panel b) that was derived from fractions 4 to 6 shown graphically in Figure 2.2. The arrows indicate annealed probes that were coincident with a single plaque on the plate.

a

b

b. Characterization of Phages

Six recombinant phage were isolated from independent initial plaques and purified, all of which proved to have the same restriction enzyme map in the insert. This 13 kb insert contained an internal Eco RI site that allowed the DNA to be cut in "halves" which were subcloned separately into pBR325. These plasmids were named pJL-001 and pJL-004 (Figure 2.4). A Northern transfer of human placental polyadenylated RNA was hybridized with [alpha-32P] dATP nick-translated inserts from these plasmids. The results showed that the insert from pJL-001 annealed to an RNA species that was about 600 nt in length while the insert from pJL-004 did not anneal to the RNA (data not shown). Taking advantage of 2 internal Bam HI sites, the insert of pJL-001 was digested and the resulting fragments subcloned separately into pBR 322. The plasmids were designated pJL-0.6, pJL-1.6 and pJL-4.6 so that the number associated with each represented the size of the insert in kb (Figure 2.4). The inserts from these plasmids were also used to probe a Northern transfer, as above, with the result that only pJL-0.6 annealed to the human placental RNA (Figure 2.5). A chicken actin cDNA probe and a yeast probe from the distal portion of the rp S10 coding gene in RPgE did not appear to anneal to the

Figure 2.4. Restriction maps of DNA inserts from a Charon 4A phage that was selected by the yeast probe fRPg (refer to Figure 2.1) and by a rat liver RNA fraction (refer to Figure 2.2). pJL-001 and pJL-004 are Eco RI fragments of 6.6 kb and 7.0 kb respectively, that comprise the human DNA insert of the selected phage. Each was cloned into pBR325. In the upper figure displaying pJL-001 and pJL-004, the scale is 1 cm = 1kb. The lower figure is not drawn exactly to scale and is presented for clarity only. pJL-001 was subcloned as two Eco RI-Bam HI and one Bam HI- Bam HI fragment which were labelled based on the size of the fragments in kb. The sequence of pJL-0.6 is presented in Figure 2.6. The restriction sites are: E. Eco RI; B. Bam HI; P. Pst I.

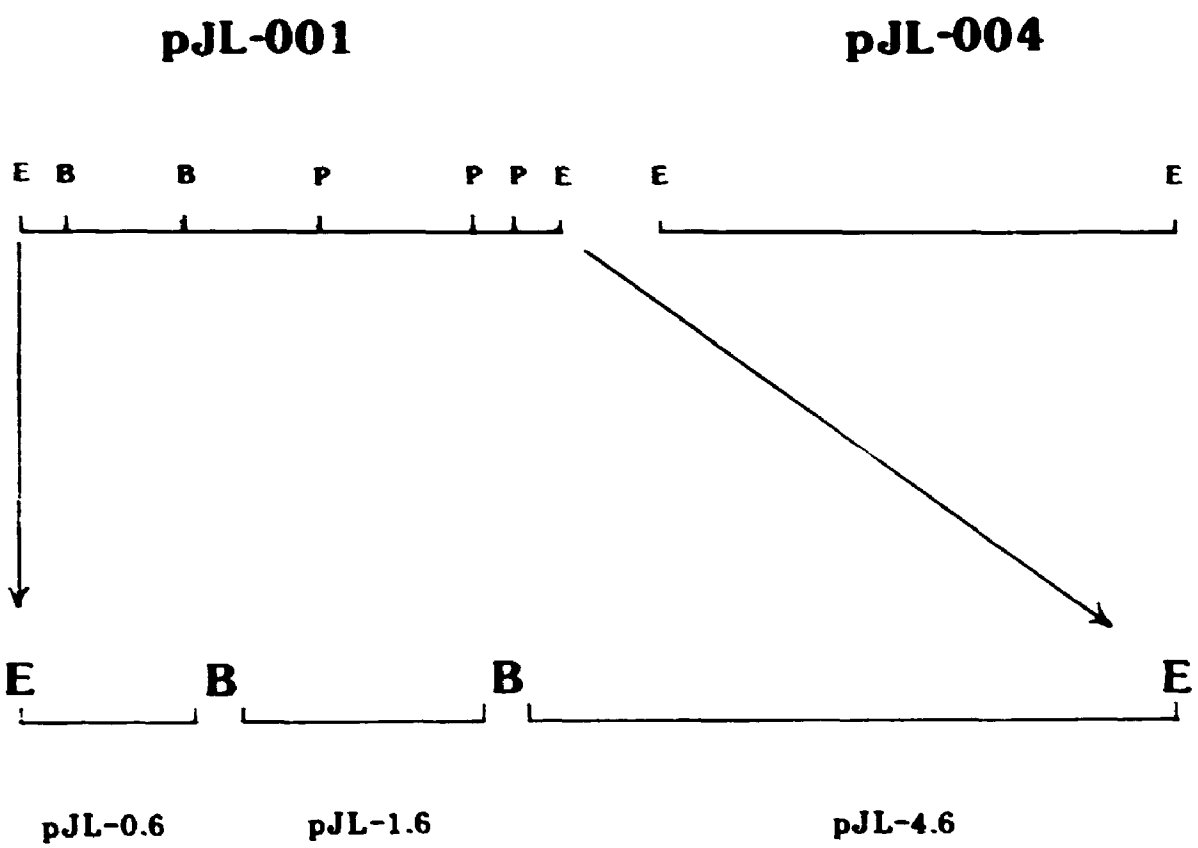

pJL-001
pJL-004
E B B P P P E E E
E B B E
pJL-0.6
pJL-1.6
pJL-4.6

Figure 2.5. Fractionation of RNA species homologous to selected human genomic DNA inserts. Samples (5 ug each) of human placental polyadenylated RNA were subjected to Northern blotting as described in section 2.2. Lane A was probed with a Hind III fragment from a chicken actin cDNA (D. T. Denhardt pers. commun. to J. B. Lott). Lane B was probed with a 650 bp Eco RI-Hind III fragment of yeast DNA that included the 3' end of the rp S10 coding gene from RPgE (refer to Figure 2.1). Lanes C, D and E were probed with the inserts from pJL-0.6, pJL-1.6 and pJL-4.6, respectively (refer to Figure 2.4).

A B C D E

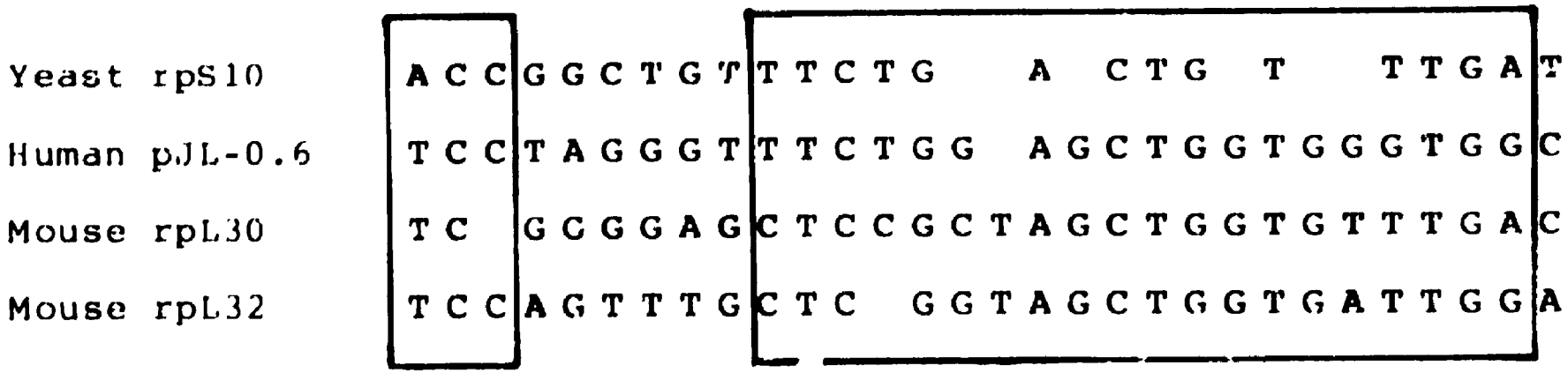
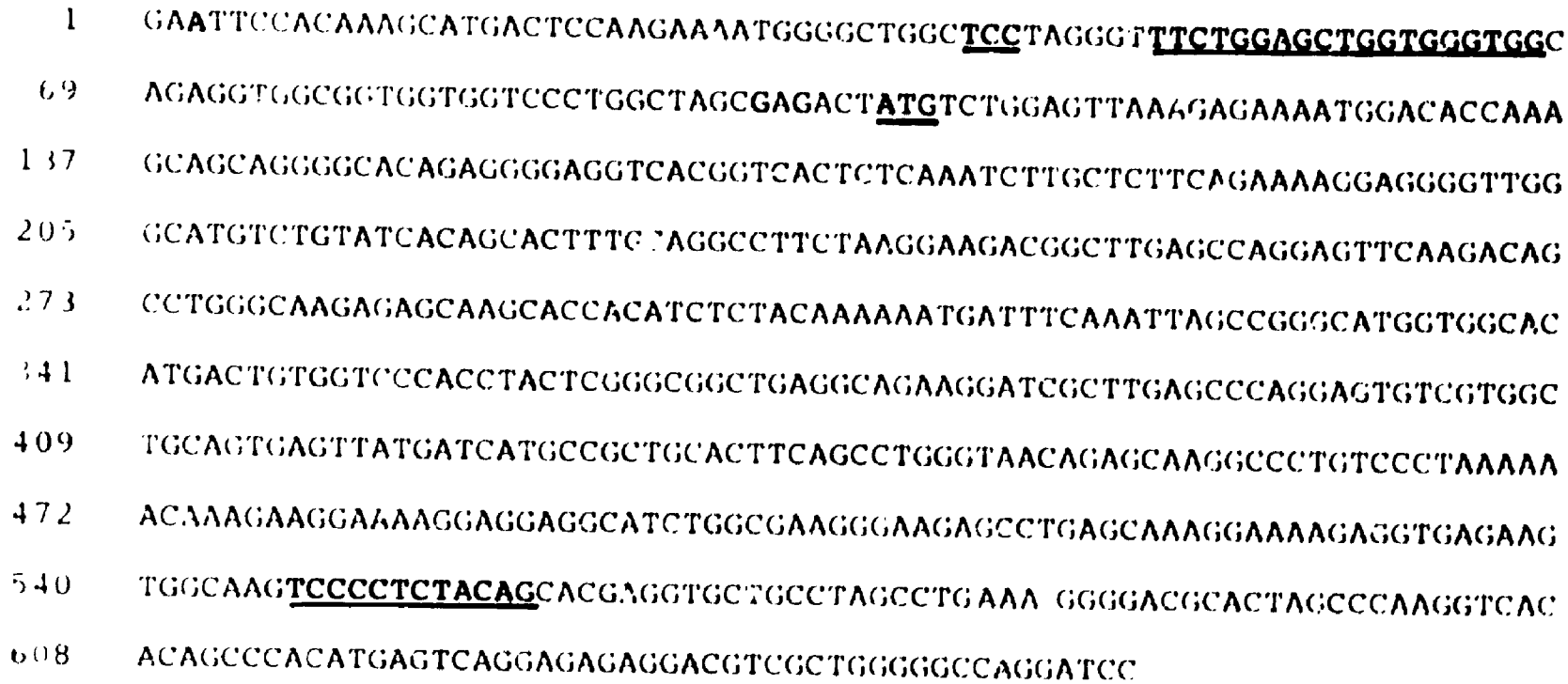

Yeast rpS10 ACCGGCTGTTTCTG A CTG T TTGAT
Human pJL-0.6 TCCTAGGGTTTCTGG AGCTGGTGGGTGGC
Mouse rpL30 TC GCGGAGCTCCGCTAGCTGGTGTTTGAC
Mouse rpL32 TCCAGTTTGCTC GGTAGCTGGTGATTGGA

Figure 2.6. Sequence of the human DNA insert from pJL-0.6. The first two groups of underlined residues represent the sequence that was found to be highly homologous to intron sequences from mouse rp genes (refer to Figure 2.7). An underlined ATG (nt 104) sequence was considered as a possible initiation codon. The underlined polypyrimidine tract followed by ACAG (nt 556) represents a consensus sequence for a 3' splice site in mammalian precusor RNA (Green, 1986). An Alu-like sequence is found at nt 194 to 465.

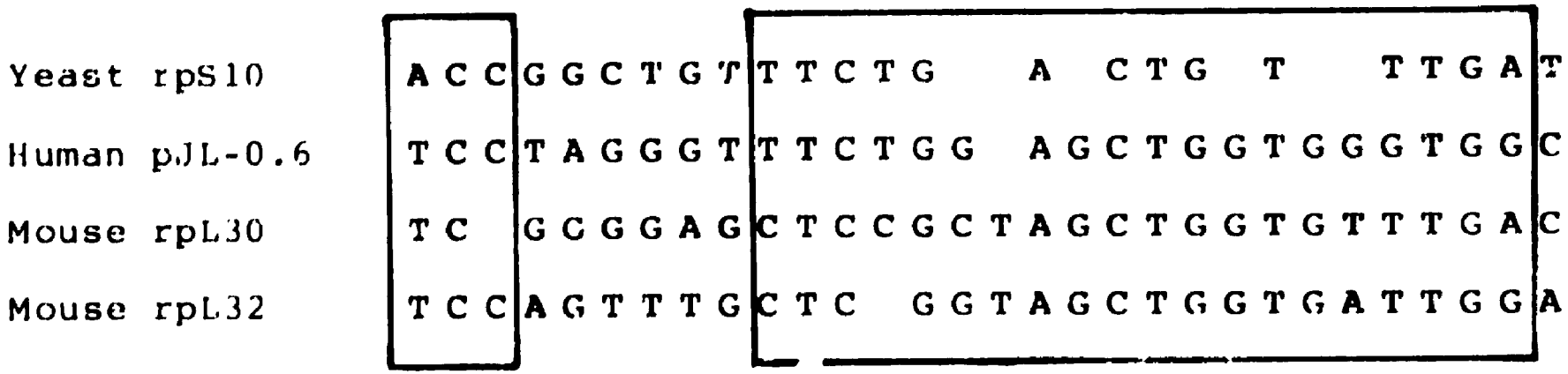

1 GAATTCCACAAAGCATGACTCCAAGAAAATGGGGCTGGCTCCTAGGGTTTCTGGAGCTGGTGGGTGGC
69 AGAGGTGGCGGTGGTGGTCCCTGGCTAGCGAGACTATGTCTGGAGTTAAAGAGAAAATGGACACCAAA
137 GCAGCAGGGGCACAGAGGGGAGGTCACGGTCACTCTCAAATCTTGCTCTTCAGAAAAGGAGGGGTTGG
205 GCATGTCTGTATCACAGCACTTTC AGGCCTTCTAAGGAAGACGGCTTGAGCCAGGAGTTCAAGACAG
273 CCTGGGCAAGAGAGCAAGCACCACATCTCTACAAAAAATGATTTCAAATTAGCCGGGCATGGTGGCAC
341 ATGACTGTGGTCCCACCTACTCGGGCGGCTGAGGCAGAAGGATCGCTTGAGCCCAGGAGTGTCGTGGC
409 TGCAGTGAGTTATGATCATGCCGCTGCACTTCAGCCTGGGTAACAGAGCAAGGCCCTGTCCCTAAAAA
472 ACAAAGAAGGAAAAGGAGGAGGCATCTGGCGAAGGGAAGAGCCTGAGCAAAGGAAAAGAGGTGAGAAG
540 TGGCAAGTCCCCTCTACAGCACGAGGTGCTGCCTAGCCTGAAA GGGGACGCACTAGCCCAAGGTCAC
608 ACAGCCCACATGAGTCAGGAGAGAGGACGTCGCTGGGGGCCAGGATCC

G-rich as are the mouse leader sequences. Second, at nucleotides 41 to 70 there is a sequence that displays striking similarity to sequences found in the first introns of the genes for mouse rp L30 and rp L32 (Figure 2.7). Third, the 50 amino acid sequence predicted by the open reading frames would generate a net basic charge of +8 if one point was allowed for a basic amino acid and one point subtracted for an acidic amino acid. This was exciting since ribosomal proteins are generally basic. Although the amino acid sequences predicted by the open reading frames were clearly not related to rp S6, the characteristics mentioned were too intriguing to disregard.

d. The Search for cDNAs Encoded by the Open Reading Frame in pJL-0.6

In order to confirm the reading frame, the genomic DNA was used to isolate the corresponding cDNA clone from a human adult liver cDNA library (Clontech, Inc.) by screening with the 0.6 kb Eco RI to Bam HI insert from pJL-0.6 (Figure 2.4). Hybridizations were conducted as before in 50% formamide and at 50 C while the washes were 4 X 15 min at 60OC in 2 X SSC and 0.1% SDS. After rescreening, 2 recombinant phage remained positive and were grown as plate lysates so that their cDNA inserts

Figure 2.7. Comparisons of DNA sequences from the introns of yeast and mouse ribosomal protein genes with a 29 nt sequence in pJL-9.6. The yeast DNA sequence is contained within the intron of one of the rp S10-coding genes of Saccharomyces carlsbergensis (Leer et al., 1982). The sequences from mouse rp L30 (Weidemann and Perry, 1984) and rp L32 (Dudov and Perry, 1984) are located in their first introns. The boxes indicate the regions of greatest homology among the sequences.


Yeast rpS10 ACCGGCTGTTTCTG A CTG T TTGAT
Human pJL-0.6 TCCTAGGGTTTCTGG AGCTGGTGGGTGGC
Mouse rpL30 TC GCGGAGCTCCGCTAGCTGGTGTTTGAC
Mouse rpL32 TCCAGTTTGCTC GGTAGCTGGTGATTGGA

could be isolated and subcloned into pBR322 for storage and into M13mp9 for sequencing. pJL-100 contained the longer of the inserts at 649 nt and pJL-5A-1A-1 had the shorter at 315 nt. A comparison of the sequences of these clones with the genomic clone, pJL-001, revealed regions of diversity but all shared a 300 nt sequence that contained only minor changes among the clones. Further study found this sequence, from nucleotide 191 to 480 in pJL-0.6, to have a high degree of homology with a member of the Alu family of repeated elements (Deininger et al., 1981). The presence of Alu sequences within introns has also been noted in the sequence of the gene encoding human rp S14 (Rhoads et al., 1986). However, the inability to select a protein coding cDNA with the genomic DNA from pJL-0.6 and the fact that the RNA species selected on the Northern transfer was approximately 600 nt clearly indicated that this clone did not encode rp S6, and probably no other ribosomal protein. The RNA species that had been identified in the Northern transfer hybridizations was probably 7S RNA since it contains an Alu-related sequence. No further investigations with the genomic DNA were conducted for this reason.

e. Direct Screening of a cDNA Library with Yeast Probes

The screening of the genomic library resulted in the isolation of a DNA fragment that certainly did not contain open reading frames that encode ribosomal protein S6. In order to avoid non-specific annealing of probe to intergenic or intron DNA, a cDNA library was probed. The human adult liver cDNA library (Clontech, Inc.) was screened with DNA fragments from the coding region of the yeast rp S10 coding gene in RPgE. Two probes, one from the 5' end of the gene and one from the 3' end of the gene, were hybridized individually at 35 C to duplicate filters for a total of four filters per plate in 50% formamide, 5 X Denhardt's solution (Denhardt, 1966), 5 X SSC, 0.1% SDS and 100 ug/ml of sonicated herring sperm DNA. Only those plaques that were positive on all four filters were picked for rescreening. One plaque met this criterion but was found on rescreening to be negative. I concluded that there was insufficient homology between the yeast probes and the human cDNAs in this library to permit detection of an rp S6 cDNA at this stringency.

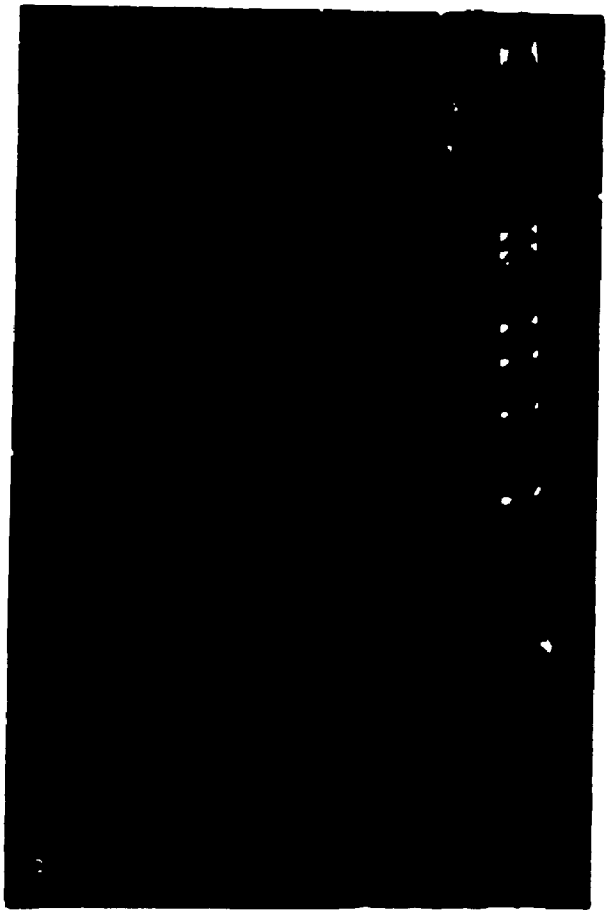
f. Cloning of the Human Homologue of Rat Ribosomal Protein S11

After the previous setbacks, another search proved to be more fruitful. When a human fibroblast cDNA plasmid library (Okayama and Berg, 1983) was screened at 42 C with a radiolabelled Bam HI to Hinf I fragment from the coding region of pRS-11-2 (Tanaka et al., 1985), the DNA from five colonies appeared to anneal to the probe. The autoradiogram in Figure 2.8 shows one such positive result. The positives were picked and streaked for single colonies. From each plate one colony was picked and grown overnight in L Broth at 37 C. Plasmid DNA was recovered by the alkaline lysis method (Maniatis et al., 1982) and then digested with Pst I and Bam HI and electrophoresed on a 6% acrylamide gel. Of the five isolates, two had restriction fragment patterns similar to that of pRS-11-2 (lanes B and C in Figure 2.9). The candidate plasmids in lanes B and C display Bam HI-Bam HI fragments of approximately 230 bp and 240 bp respectively. Since these represent fragments from the 3' end of the cDNA inserts the differences in electrophoretic mobility reflect differences in the lengths of the cloned polyadenylate tails. Lanes B and C also display Pst I-Bam HI fragments of 400 bp and 380 bp respectively which represent

Figure 2.8. Autoradiograph of a nitrocellulose filter colony lift screened with a rat rp S11 cDNA probe. An aliquot of a human fibroblast cDNA plasmid library (Okayama and Berg, 1983) was screened with a 208 bp Bam HI-Hinf I DNA fragment from the rat rp S11-coding cDNA insert in pRS-11-2 (Tanaka et al., 1985). The arrows indicate a positive colony that was later identified as containing a plasmid (pHS11) with a full length cDNA insert encoding human rp S11 (see text).

Figure 2.9. Electrophoretic analysis of digests of plasmids recovered from colonies selected by a rat rp S11 probe. DNAs were digested with the restriction enzymes enumerated below and the products resolved on a neutral 6% polyacrylamide gel. DNA fragments were visualized by staining with ethidium bromide. Lanes A to E represent Pst I and Bam HI digests of candidate plasmids isolated from a human fibroblast cDNA library (Okayama and Berg, 1983). Lane F is a Pst I and Bam HI digest of pRS-11-2, a plasmid containing a cDNA insert encoding rat rp S11. Lane G shows markers of known size derived from a Hinf I digest of pBR322.

A B C D E F G



1631

517

506

396

344

298

220

154

75

differences in the 5' extent of each insert. The identification of bands as Bam H1-Bam H1 or Pst 1-Bam H1 was confirmed by other digestions and analyses including digestion with Bam HI alone(data not shown). A Bam HI restriction enzyme site in the vector that is only 30 bp distal to the 3' end of the cDNA insert facilitated the assignment of the fragment containing the polyadenylated tail as the Bam HI-Bam HI fragment (see also Figure 2.10).

The plasmid in lane B of Figure 2.9 (subsequently called pHS11) was picked, the insert isolated and subcloned into M13 based vectors and sequenced. The sequencing strategy is shown in Figure 2.10 and the DNA sequence is shown in Figure 2.11 (taken from Lott and Mackie, 1988). The 5' end of the insert in lane C was also sequenced to confirm that it was shorter than the insert in lane B. The human rp S11 cloned cDNA consists of 15 nt of the 5' leader, the entire coding sequence and all 51 nt of the 3' untranslated region. The predicted amino acid sequence of 158 residues is identical to that of rat rp S11. The nucleotide sequence in the coding region differs, however, from that in rat in the first position in two codons and in the third position in 44 codons.

Figure 2.10. Sequencing strategy of the cDNA insert in pHS11. This plasmid was isolated from a human fibroblast cDNA library (Okayama and Berg, 1983) as described in the text and contains homopolymeric (GGG) tails at each end (not shown at the 3' end). The arrows indicate the extent to which the sequence was determined from individual phage. The restriction sites are: P, Pst I; H, Hae III; R, Rsa I; S, Sau 3a; B, Bam HI.

pHS-11

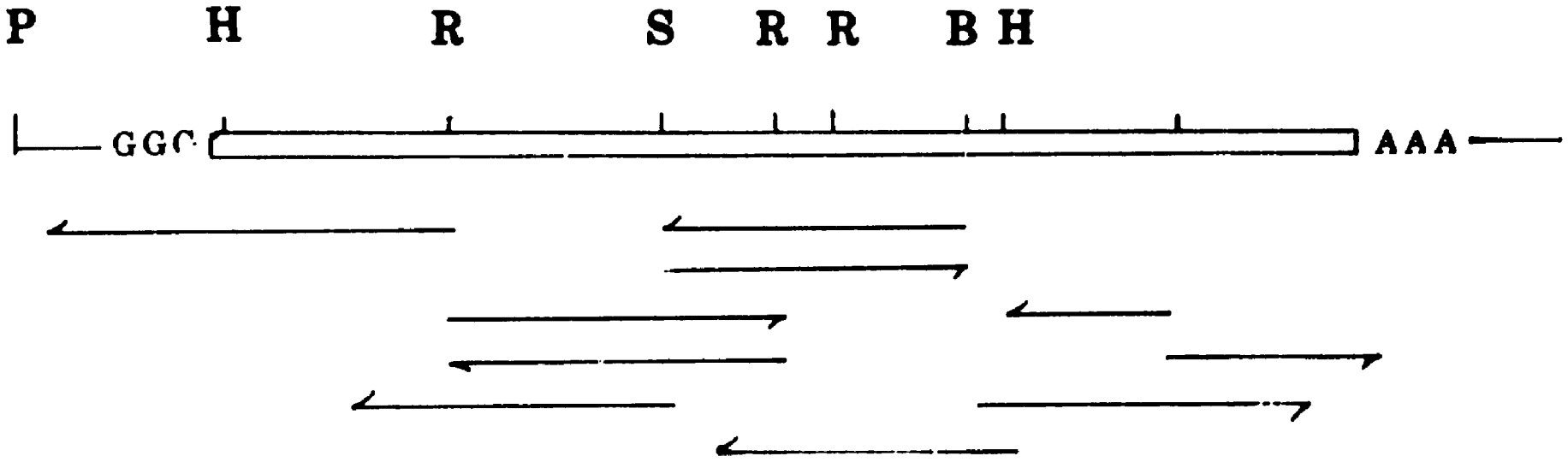


Figure 2.11. The cDNA sequence and predicted amino acid sequence, as inferred by the insert in pHS11, of human ribosomal protein S11 . Numbering is from the first residue of the cDNA insert. The designated nucleotide is aligned with the last digit of each number.

```
                  Met Ala Asp Ile Gln Thr Glu Arg Ala Tyr Gln Lys Gln Pro
CAGGCGGCCGGGAAG ATG GCG GAC ATT CAG ACT GAG CGT GCC TAC CAA AAG CAG CCG
  3
Thr Ile Phe Gln Asn Lys Lys Arg Val Leu Leu Gly Glu Thr Gly Lys Glu Lys
ACC ATC TTT CAA AAC AAG AAG AGG GTC CTG CTG GGA GAA ACT GGC AAG GAG AAG
 60
Leu Pro Arg Tyr Tyr Lys Asn Ile Gly Leu Gly Phe Lys Thr Pro Lys Glu Ala
CTC CCG CGG TAC TAC AAG AAC ATC GGT CTG GGC TTC AAG ACA CCC AAG GAG GCT
114
Ile Glu Gly Thr Tyr Ile Asp Lys Lys Cys Pro Phe Thr Gly Asn Val Ser Ile
ATT GAG GGC ACC TAC ATT GAC AAG AAA TGC CCC TTC ACT GGT AAT GTG TCC ATT
168
Arg Gly Arg Ile Leu Ser Gly Val Val Thr Lys Met Lys Met Gln Arg Thr Ile
CGA GGG CGG ATC CTC TCT GGC GTG GTG ACC AAG ATG AAG ATG CAG AGG ACC ATT
222
Val Ile Arg Arg Asp Tyr Leu His Tyr Ile Arg Lys Tyr Asn Arg Phe Glu Lys
GTC ATC CGC CGA GAC TAT CTG CAC TAC ATC CGC AAG TAC AAC CGC TTC GAG AAG
276
Arg His Lys Asn Met Ser Val His Leu Ser Pro Cys Phe Arg Asp Val Gln Ile
CGC CAC AAG AAC ATG TCT GTA CAC CTG TCC CCC TGC TTC AGG GAC GTC CAG ATC
330
Gly Asp Ile Val Thr Val Gly Glu Cys Arg Pro Leu Ser Lys Thr Val Arg Phe
GGT GAC ATC GTC ACA GTG GGC GAG TGC CGG CCT CTG AGC AAG ACA GTG CGC TTC
384
Asn Val Leu Lys Val Thr Lys Ala Ala Gly Thr Lys Lys Gln Phe Gln Lys Phe
AAC GTG CTC AAC GTC ACC AAG GCT GCC GGC ACC AAG AAG CAG TTC CAG AAG TTC
438

TGAGGCTGGACATTCGGCCCGCTCCCACAATGAAATA*AGTTATTTTCTATTCC
492
```

2.4 Discussion

The value of cross-hybridization as a means of locating a particular DNA sequence in one species by using an analogous sequence from a different species has been discussed above (refer to the Introduction). While this method is often successful it does have limitations that must be considered when designing cross-hybridization experiments.

As detailed in Chapter 1 there can be extensive homology in the amino acid sequences of ribosomal proteins from closely related species. Even where two proteins from different species display a high percentage of identical amino acids there can still be substantial differences in the nucleotide sequence. Rat rp S11 and human rp S11 exhibit 100% identity of amino acids but only 90% identity at the nucleotide level. These changes are due to the degeneracy of the genetic code and can reflect any in codon usage bias an organism or tissue may display. Examination of the first position changes between the coding regions of the rat and human rp S11 cDNA sequences shows that one of these is a CGT to AGG change while the other is AGG to CGG. Since these codons all encode arginine and since one first position change is the reverse of the other there does not appear to be a

discernible trend in codon usage from rat to human for this protein. The sample size represented by these data is too small to allow any generalizations concerning codon changes between other human and rodent ribosomal protein cDNAs.

When the target tissue and the probe are from species that are not closely related there are usually amino acid differences between analogous proteins and therefore codon changes leading to divergence of their nucleotide sequences. For example, the N termini of rat rp S6 and yeast rp S10 are 60% homologous in their amino acid sequences. One should expect significant differences in nucleotide sequence. Extensive mismatches in nucleotide sequence could account for the failure of the yeast rp S10 encoding DNA to select human rp S6 coding DNA from the human genomic library. There are, however, stretches of 15 amino acids from residues 54 to 68 in yeast rp S10 (Leer et al., 1982) and 54 to 68 in human rp S6 (Lott and Mackie, 1988) that are identical (see Figure 3.5) and whose nucleotide sequences differ in only four positions. Within this region there are 20 nt from +166 to +185 that are identical. Although this degree of homology may be sufficient to identify the desired human clones, this 45 nt sequence comprises just 7% of the 3' probe from the gene encoding yeast rp S10 and may not

have been labelled sufficiently to be detected over background.

While it is likely that the yeast probe was the weak link in the original experimental design, it is also possible that the aliquot of the human genomic library that was screened may not have contained representative amounts of the desired clones. The problem anticipated at the beginning of this project was in sorting out the false positives from the desired clones. In the case of these experiments one should expect that the yeast DNA probe would anneal to several clones since it contains some intergenic and intron DNA and has just 60% nuclectide homology between the coding region it carries and its analogous human sequence. Instead, one clone was selected six times implying that it is highly, and thus disproportionately, represented in this sample of the genomic library. The library was made by Lawn et al. (1978) and has since been amplified and passed to other laboratories. Amplification of phage and plasmid libraries may result in selection of particular clones due to differences in the growth characteristics conferred on individual phage by their genomic or cDNA inserts.

The presence of an Alu sequence in the DNA insert is not surprising since 94% of all the genomic recombinants have such sequences (Hwu et al., 1986) As stated

earlier, the third and fourth introns of the gene encoding human ritosomal protein S14 contain Alu sequences (Rhoads et al., 1986). The genomic DNA comprising the insert of pJL-0.6 appears to have some of the structural properties of eukaryotic genes but the authenticity of these elements is not supported by its failure to select bona fide cDNAs. One must conclude that pJL-0.6 does not carry protein coding DNA expressed in the liver.

On the other hand, pHS11 contains a cDNA insert encoding human rp S11. The human rp S11 and rat rp S11 amino acid sequences are identical and therefore share the same physical characteristics (hydropathicity, helical content etc., already discussed in Tanaka et al., 1985). The nucleotide sequences encoding these two proteins display polymorphism in 10% of their positions demonstrating nucleotide sequence drift that does not affect amino acid sequence.

CHAPTER 3

ISOLATION OF HUMAN RP S6 cDNAs USING MIXED OLIGONUCLEOTIDE PROBES

3.1 Introduction

The strategy of screening cDNA or genomic libraries with oligonucleotide probes whose sequences are based on amino acid sequence data has been used by others quite successfully (for a review see Itakura et al., 1984). In the first reported application of this technique Suggs et al. (1981) detected a human Beta-2-microglobulin cDNA in a screen of 535 bacterial plasmid clones. The heterologous probes were designed to encode two regions of the known amino acid sequence of human Beta-2-microglobulin. The first region, residues 95 to 99, required two series of 15 mers to account for all possible combinations of codons due to arginine at residue 97. The heterologous probe in which arginine was represented by AGA or AGG exhibited 8-fold degeneracy whereas the probe in which arginine was represented by CGN, where N is all four nucleotides, displayed 16-fold degeneracy. The second region, residues 75 to 78, was encoded by an 8-fold degenerate 11 mer. The 15 mer having the CGN codons annealed to a cloned cDNA that was shown to encode human Beta-2-microglobulin. It is not surprising that the 15 mer containing AGG or AGA did not anneal. On the other hand, the 11 mer did not anneal to the human Beta-2-microglobulin cDNA even though it

contained a sequence that was perfectly complimentary to the cDNA sequence. The 11 mer may have been too short to anneal specifically to the desired cloned cDNA consistent with observations (Derynck et al., 1984) suggesting that oligodeoxynucleotide probe length may be a limiting factor in the usefulness of some probes.

The human insulin-like growth factor (IGF-1) gene was isolated using a 103 bp probe that was based on the known amino acid sequence of the IGF-1 beta chain (Ullrich et al., 1984). The double stranded probe contained an arbitrary nucleotide at each degenerate position. Overall homology between the probe and the target sequence proved to be 57%; within the probe, however, there was an 81 bp segment that displayed 80% homology with the IGF-1 gene coding sequence.

A human factor IX cDNA clone was isolated using a single-stranded probe that was comprised of two consecutive 26 mers thus forming a unique 52 mer (Jaye et al., 1983). As these authors' original intention was to first isolate a bovine factor IX cDNA clone, the probe sequence was based on bovine amino acid data and nucleotides were chosen at degenerate positions according to codon usage in known bovine DNA coding sequences. This probe was also used to screen 104 colonies of a human liver cDNA library in addition to bovine cDNA libraries. The probe annealed to only one colony in the human

library which was found to contain a cloned cDNA encoding human factor IX. The overall homology between the probe and the human target was relatively high, 85%, but the longest stretch of perfect homology was only 14 nt.

In an effort to isolate the gene coding for human transforming growth factor alpha (TGF-alpha), Derynck et al. (1984) used several synthetic probes whose sequences were based on partial amino acid sequence data. A 41 mer and a 48 mer were designed such that their 3' ends were complimentary over 15 nt so that they could also be annealed and extended to form a double-stranded 74 bp probe. The choices of nucleotides at degenerate positions were based on codon usage in human sequences (Grantham et al., 1981) and avoidance of CpG dinucleotides which are uncommon in human DNA (Bird, 1980). In addition, this group synthesized two sets of non-overlapping 14 mers, referred to as pool 1 and pool 2, each having 16 fold degeneracy. A cDNA library derived from a cell line that produces TGF-alpha was screened without success. After using all the probes to search the cDNA library, a human genomic library (Lawn et al., 1978) was screened with the single-stranded 41 mer or double-stranded 74 mer with a better outcome. Apparent positives were rescreened on dot blots with the 41 mer, 48 mer, 74 mer and both pools of 14 mers. None hybridized to the 48 mer or to pool 1. Five recombinant phage from the library remained positive

after the rescreen and all were found to share the same 180 bp Sau 3A fragment which was in turn used to isolate a TGF-alpha cDNA clone. One of the 14 mers of pool 2 was perfectly homologous to part of the TGF-alpha coding sequence. Oligodeoxynucleotides from pool 1 did not anneal to this cDNA as their sequences were based on an amino acid sequence which was later shown to be inaccurate.

The examples cited above show that oligodeoxynucleotide probes can be useful in identifying DNA sequences which encode specific proteins when some amino acid data for the desired protein is available. As detailed above, the length of the probe can be a determinant of success. This is demonstrated by the failure of a perfectly complimentary probe that was too short to anneal and by the success of probes that are long enough to allow inclusion of a few incorrect nucleotides at degenerate positions. The length of heterologous probes will usually be determined by the degree of acceptable degeneracy; therefore, probes based on amino acid sequences that display as little degeneracy as possible are most desirable. Finally, errors in protein sequence compromise the usefulness of any degenerate probe, but most dramatically of relatively short probes.

3.2 Materials and Methods

a. Oligodeoxynucleotide Probes

Oligodeoxynucleotides were synthesized on an Applied Biosystems Model 380A Synthesizer employing control pore glass solid phase method developed by Matteucci and Caruthers (1981) combined with proton activated nucleoside phosphoramidites developed by Beaucage and Caruthers (1981). Crude products were separated by preparative polyacrylamide gel electrophoresis and visualized by ultraviolet light (UV shadowing). Gel slices containing the desired oligodeoxyrucleotides were excised. The oligodeoxynucleotides were eluted as described in section 2.2k but without added tRNA.

Oligodeoxynucleotide probes were radiolabelled in a manner similar to that for end labelling RNA (see section 2.2e) and free nucleotides were removed on Sephadex G-25 columns.

b. Human Placental cDNA Library

An aliquot containing 40,000 PFU of a human placental cDNA library carried by the vector lambda gt10 (Ullrich et al., 1984) was obtained from Dr. J. Riordan. The bacterial host was *E. coli* C600 rk- mk+ hfl and was

grown on LB agar.

c. Screening Human cDNA Libraries

A portion of a human fibroblast cDNA library (Okayama and Berg, 1983; see section 2.2a) was plated on twenty-two 150 mm plates at a density of 1500 colonies per plate. Colony lifts were performed with nitrocellulose filters which were then denatured, neutralized and dried in a vacuum oven. The filters were prehybridized for 6 h at 25 C in a buffer containing 6 X SSC, 1 X Denhardt's solution (Denhardt, 1966), 0.05% SDS, 6 mM EDTA and sonicated herring sperm DNA at 100 ug/ml. The filters were then placed in fresh buffer to which 32P-end-labelled probe was added at 500,000 cpm/ml. Hybridization continued for 16 h at 25 C followed by three 10 min washes at 25 C and then one 10 min wash at 37 C in 2 X SSC containing 0.1% SDS. Dried filters were exposed to film for 24 h after which the apparent positives were picked, replated and rescreened.

Hybridizations were performed with nick translated probes at 42 C in 50% formamide, 5 X SSC, 5 X Denhardt's solution (Denhardt, 1966), 0.1% SDS, 10 mM PIPES (pH 6.4) and 250 ug/ml sonicated herring sperm DNA for 6-24 h. Hybridizations in fresh buffer continued for 16-24 h and were followed by four washes with 2 X SSC containg 0.1%

SDS for 15 min each at 42 C. Apparent positives were replated and rescreened.

d. Preparation and Analysis of Cellular DNA and RNA

Hela S3 cells were obtained from Dr. I. G. Walker and were grown in alpha-modified Eagles' medium containing 10% Fetal Bovine Serum, and the antibiotics penicillin, streptomycin and tylocine. Total RNA was isolated from Hela S3 cells with guanidine-HCl (Strohman et al., 1977). Polyadenylated RNA was selected using affinity chromatography on a 1 ml column of oligo(dT)-cellulose (Maniatis et al., 1982). RNA was denatured with glyoxal, electrophoresed at 5 ug/lane in a 1.5% agarose gel and then transferred to a Biodyne A nylon membrane (Pall Inc.) (Mackie, 1986). RNA was fixed to the support by UV irradiation (Khandjian, 1987). Blots were prehybridized for 6 h at 42 C then hybridized with 250,000 cpm/ml of the appropriate probe for 16 h at the same temperature (as described by Mackie, 1986) and then washed for 1 h with four changes of 2 X SSC containing 0.1% SDS at 48 C (Maniatis et al., 1982). Blots were exposed to Kodak XAR film for 6 to 20 h. Internal fragments of pHS6-1, pHS11 and yeast rp S10 DNA were labelled with [alpha-32P]dCTP using the random priming method (Feinberg and Vogelstein, 1983).

High molecular weight human T cell DNA, a gift from Dr. J. K. Ball, was prepared as described (Canaani and Aaronson, 1979), digested with restriction enzymes and then electrophoresed on a 0.8% agarose gel. Each sample contained 15 ug of human genomic DNA. A second set of digestions was performed under similar conditions with the exception that each sample also included 1 ug of lambda DNA. After electrophoresis the second set of lanes was stained with ethidium bromide and visualized under ultraviolet light to determine if the lambda DNA had been fully digested (data not shown). This procedure was employed to ensure that all of the DNA in the sample had been fully digested. Digested DNAs from the first set of samples were transferred to nylon with 10 X SSC, fixed and annealed with probes as described above. The probes consisted of nick-translated cDNA inserts from pHS6-1, pHS6-2A and pHS6-2B labelled with [alpha-32P]dCTP and [alpha-32P]dGTP.

e. Slot Blot Analysis of RNA

Blots of polyadenylated RNA were hybridized with radiolabelled probes and then analyzed by densitometry. RNA was applied to the Biodyne A nylon membrane (Pall Inc.) with a slot blot manifold (Schleicher and Schuell, Inc.) such that the first slot in each row received 2.5

ug of RNA, the second received 0.525 ug RNA and the third received 0.156 ug RNA with all RNA samples dissolved in 10 X SSC. The first row in each blot consisted only of 10 X SSC and no RNA, the second row contained total E. coli RNA, the third had polyadenylated RNA from Hela S3 cells and the fourth row contained Hela S3 RNA minus the polyadenylated fraction. RNA was fixed to the support by UV irradiation (Khandjian, 1987) and then the blots were prehybridized at 47 C in 50% formamide, 5 X SSC, 5 X Denhardt's solution (Denhardt, 1966), 0.1% SDS, 10 mM PIPES (pH 6.4) and 250 ug/ml of sonicated herring sperm DNA for 6 h. Probes were made by the random priming method with [alpha-32P]dCTP on 250 bp Hae III fragments from the coding region of pHS11 and pHS6-1. The probes were added to the appropriate filters at 100,000 cpm/ml and hybridization continued for 22 h. Filters were washed for 1 h at 45 C four times in 2 X SSC containing 0.1% SDS and then placed on Kodak XAR film for 24 h. The resultant radiographs were scanned by an LKB 2222-010 Ultro scan XL Laser Densitometer and the data from those slots in the linear range were compared (refer to Figure 3.12a).

f. Computational Methods

A comparison of amino acid homology between human rp S6 (Lott and Mackie, 1988) and yeast rp S10 (Leer et al.,

1982; 1985) was performed with a Zenith personal computer. The two dimensional plot compares the two amino acid residues beginning at amino acid 1 of each such that only a perfect match of 4 amino acids between each sequence is scored (Lagrimini et al., 1984).

Hydropathic plots of the amino acid sequences of human rp S6 and yeast rp S10 were constructed with a program that uses a 9 element running average similar to a Kyte and Doolittle (1982) plot.

3.3 Results

a. Rationale and Design of Probes

Lacking amino acid sequence data for human rp S6, I compared the partial amino acid sequence of rat liver rp S6 (Wool, 1979; Wettenhall and Morgan, 1984) to the complete sequence of yeast rp S10 (Leer et al., 1982; 1985) to identify regions of conserved sequence that could be used to design probes with as little degeneracy as possible.

Two sets of non-overlapping probe were designed and synthesized in the expectation that only bona fide rp S6 cDNAs would anneal to one member of each set of probes. The sequence of one such set of probes ("N-terminal probe") was based on amino acids 17 to 21 in both yeast rp S10 and rat rp S6 choosing valine at position 18 as in rat rather than isoleucine as is found in yeast (refer to Figure 3.1). A second set of probes ("C-terminal probes") was designed from the least degenerate region of the C terminus of the published rat rp S6 amino acid sequence (Wettenhall and Morgan, 1984). The chosen region is part of a 10 amino acid C terminal sequence present in rat liver rp S6 but not found in yeast rp S10. Two mixed oligodeoxynucleotide probes were synthesized that encoded the last five amino acids of rat liver rp S6 and the

Figure 3.1. Sequences of the N and C terminal mixed oligodeoxynucleotide probes. The upper figure shows the aa sequences of residues 14 - 24 near the amino termini of yeast rpS10 (Leer et al., 1982; 1985), rat rpS6 (Wool, 1979) and human rpS6. The leucine at residue 16 in the rat rpS6 sequence is now known to be isoleucine (Chan and Wool, 1983). The set of probes based on the rat and yeast aa sequences is shown on line 4 and details of its synthesis are given in section 3.2. Two nucleotides at one position indicates that both were introduced during synthesis in equimolar quantities. The actual nt sequence of the human rpS6-coding cDNA is given below the nt sequence of the probe.

The lower portion of the figure shows the aa sequence of the C terminus of rat rpS6 (Wettenhall and Morgan, 1984), and for comparison, the inferred aa sequence of human rpS6 below it. The nt sequences of the two C-terminal probes and the actual human rpS6-coding cDNA (see Fig. 3.6) are provided. The human sequence is taken from Lott and Mackie (1988b).

N TERMINAL PROBES

Yeast S10	.. lys thr phe glu ile asp asp glu his arg ile ..
Rat S6	.. lys leu leu glu val asp asp glu arg lys leu ..
Human S6	.. lys leu ile glu val asp asp glu arg lys leu ..
Probe	$GA\frac{A}{G}$ GTI $GA\frac{C}{T}$ $GA\frac{C}{T}$ GA
Human S6 sequence	... GAA GTG GAC GAT GAA ...

C TERMINAL PROBES

Rat S6	.. ser thr ser lys ser glu glu ser gln lys stop
Human S6	.. ser thr ser lys ser glu ser ser gln lys stop
Probes	$GA\frac{A}{G}$ $GA\frac{A}{G}$ $AG\frac{C}{T}$ $CA\frac{A}{C}$ $AA\frac{A}{G}$ T
	$GA\frac{A}{G}$ $GA\frac{A}{G}$ TCI $CA\frac{A}{G}$ $AA\frac{A}{G}$ T
Human S6 sequence	... GAA TCC AGT CAG AAA TAG ...

first T of the stop codon. Two sets of probes were necessary to accommodate all six serine codons. In an effort to limit the degeneracy of the probes, inosine was selected for positions where all four bases were possible since between 25 C and 50 C inosine is the least discriminatory base (Martin et al., 1985; Figure 3.1).

b. Screening the cDNA Libraries with Oligodeoxynucleotide Probes for Ribosomal Protein S6 Coding Sequences

An initial attempt at probing a human fibroblast cDNA library (Okayama and Berg, 1983) demonstrated that the C-terminal probes displayed unanticipated complimentarity to the vector resulting in a high background. In addition, it was later discovered that the most distal glutamate in rat rp S6 reported by Wettenhall and Morgan, (1984) is replaced by serine in our sequence (refer to Figure 3.1). This unexpected difference may account for the lack of specificity of the C-terminal probes. The shorter N-terminal probe demonstrated high specificity and annealed to only one colony in 33,000 (see Figure 3.2). This colony was purified by replating and reprobing at increasing dilutions a total of 4 screenings).

A 550 bp Pst 1 fragment of the plasmid (pHS6-1) purified from this colony was subcloned into M13mp9 and

Figure 3.2. Autoradiograph of a nitrocellulose filter colony lift screened with the N terminal probe. An aliquot of a human fibroblast plasmid cDNA library (Okayama and Berg, 1983) was screened with a 14 mer oligodeoxynucleotide probe whose sequence was based on N terminal amino acids in rat rp S6 and yeast rp S10 (refer to section 3.3a). To accommodate the degeneracy of the genetic code the probe actually consisted of 32 different oligonucleotide sequences. The arrow indicates a positive signal associated with a colony that was later replated, rescreened, purified, and determined to contain a plasmid carrying a cDNA insert encoding human rp S6 (pHS6-1). The apparent positive that is 1 cm above the arrowhead was caused by the probe annealing non-specifically to bacterial debris. Three reference points appear on the radiograph.

was partially sequenced. The results demonstrated the presence of an open reading frame on one strand of the cloned DNA. The predicted amino acid sequence of this open reading frame was highly homologous to a portion of the sequence of yeast rp S10 (Figure 3.3). Subfragments of this human cDNA insert were isolated both for cloning into sequencing vectors and for labelling as probes to rescreen the library. Further screening yielded shorter clones but none longer.

After the sequence of the entire cDNA insert was completed, it was apparent that pHS6-1 was not full length. The N-terminal sequence predicted by the cDNA insert of pHS6-1 displayed a high degree of homology with the N-terminal sequence of yeast S10 but was shorter by 6 amino acids. Restriction maps of other shorter clones isolated from this library indicated that none were longer at the 5' end.

In order to find full-length clones we decided to screen a human placental cDNA library in lambda gt10 with the 550 bp Pst I fragment from pHS6-1. This library had previously yielded a full length cDNA (5 kb) encoding human epidermal growth factor receptor (Ullrich et al., 1984) . Four plaques remained positive after rescreening and their inserts were subcloned into pSP65. Three of these were shorter than pHS6-1 and one was longer (pHS6-2B). One of the shorter clones (pHS6-2A) also proved to

Figure 3.3. Comparison of nucleotide and inferred amino acid sequences between an open reading frame in pHS6-1 and a highly homologous region from a yeast gene encoding rp S10 (Leer et al., 1982). The region of comparison starts at amino acid 143 of yeast rp S10. Only those yeast nucleotides and amino acids that differ from the human sequence are shown.

```
                  143
                  lys leu phe asn leu ser lys glu asp asp val
pHS6-1            AAA CTT TTC AAT CTC TCT AAA GAA GAT GAT GTC
Yeast S10         --G T-C --- GG- T-G --C --G --- --- --C --T

                      phe     gly
                  143

                  arg gln tyr val val arg lys pro leu asn lys

                  CGC CAG TAT GTT GTA AGA AAG CCC TTA AAT AAA
                  --T G-T -TC --C A-C --- -GA GAA G-C -CC --G

                      asp phe     ile     arg glu val thr

                  glu gly lys lys pro arg thr lys ala pro lys

                  GAA GGT AAG AAA CCT AGG ACC AAA GCA CCC AAG
                  -GT -AA --- -CT TAC     --- --G --T --A ---

                  gly glu     thr tyr

                  ile glu arg leu val thr pro arg val leu gln

                  ATT CAG CGT CTT GTT ACT CCA CGT GTC CTG CAG
                  --C --A A-A T-G --- --- T-T -AA AGA T-- --A

                      gln                 ser gln arg
```

be of interest and was sequenced as well (refer to Figure 3.4).

c. Nucleotide sequence of human rp S6-coding region

The combined sequence data from the three cDNA clones characterized above accounts for the entire coding sequence (249 aa), all 39 bp of the 3' untranslated region and 27 bp of the 5' untranslated region (refer to Figures 3.5 and 3.6). The complete nucleotide sequence of the cloned cDNA in pHS6-1 includes the 3' untranslated region and all but 18 nt of the coding sequence. The cloned cDNA in pHS6-2B is longer than pHS6-1 by 12 nt at the 5' end, but is otherwise identical even though isolated from a library prepared from a different tissue. The size and the extent of the 390 bp insert in pHS6-2A suggests that it may have been primed at a sequence of seven adenines (nucleotides 342 to 348 in Figure 3.6). This insert also contains an inversion and a deletion so that nucleotides +63 to -27 are followed by +93 to +348. In the regions in which pHS6-2A and pHS6-2B overlap, the sequences are identical. The most interesting structural feature of rp S6 mRNA is a hairpin which can be created with nucleotides -2 to +1 forming the loop and nucleotides -7 to -3 base pairing with nucleotides +2 to +6. An internal loop can be formed with nucleotides -8

Figure 3.4. Sequencing strategy of rpS6-coding cDNA inserts. The diagram represents the alignment of rpS6-coding cDNA inserts obtained from two different libraries (see section 3.3). pHS6-1 was isolated from a human fibroblast cDNA library (Okayama and Berg, 1983) and contains homopolymeric (GGG) tails at each end. pHS6-2A and pHS6-2B were isolated from a human placental cDNA library (Ullrich et al., 1984). The cDNA insert of pHS6-2A is represented as it was isolated and sequenced in line 1. In line 2 the insert is shown as it aligns with pHS6-2B (line 3) and pHS6-1 (line 4) with the hatched area referring to the inversion explained in the text. The arrows indicate the direction and extent to which the sequence was determined from individual phage sequences. The restriction sites are: E. Eco RI; H. Hae III; P. Pst I; S. Sau 3A.

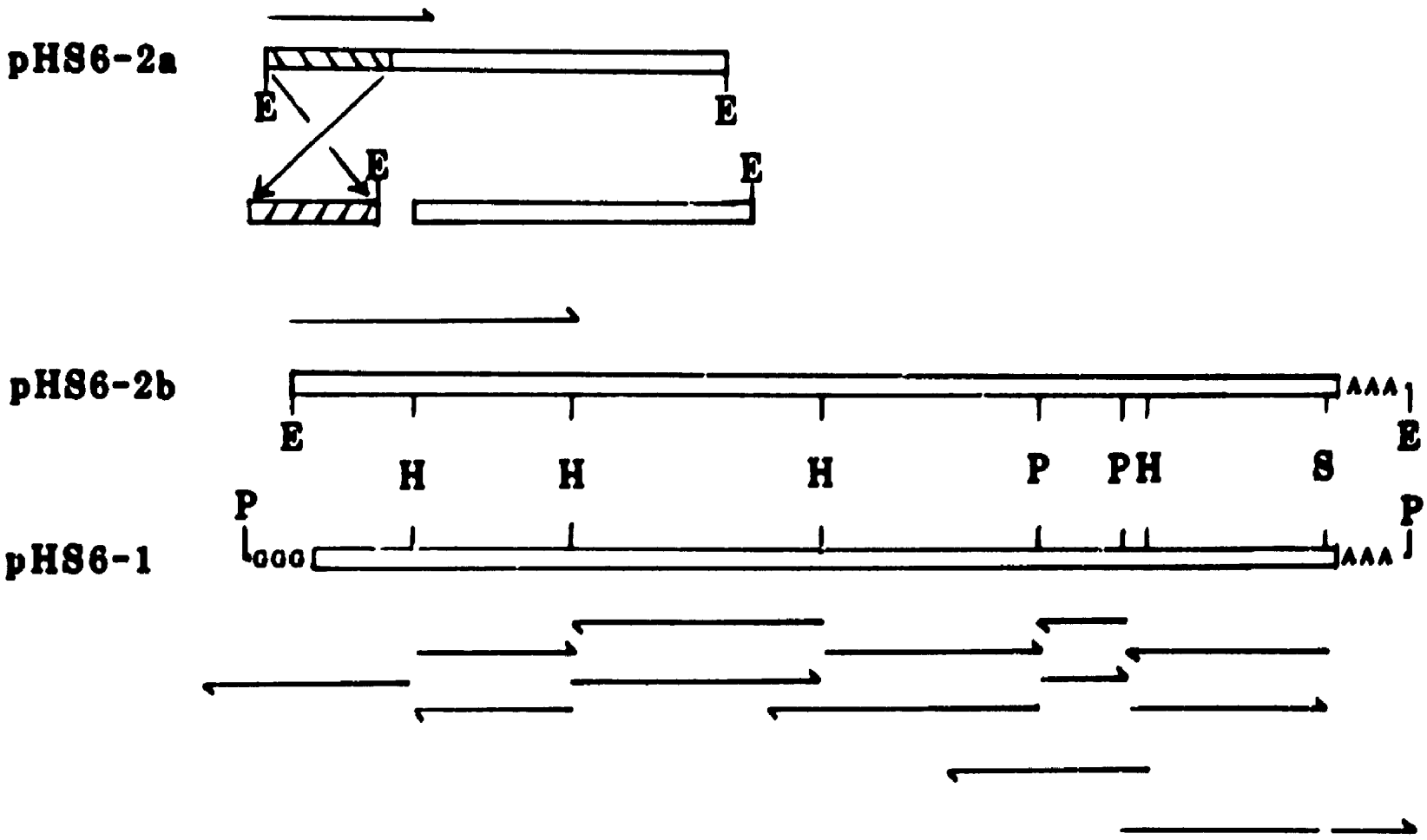

pHS6-2a
E
E
E
E
pHS6-2b
E
H
H
H
P
PH
S
E
P
pHS6-1
P

Figure 3.5. Sample of a sequencing gel. The autoradiograph depicts a sequence ladder generated with [35S]dATP employing the dideoxy chain termination method (Biggin et al., 1983). The nucleotide represented in each lane is indicated at the bottom of the autoradiograph. The ladder on the right corresponds to the sequence of a human rp S6 cDNA, in a 5' to 3' direction, about and including nucleotides 463 to 467 which are labelled (refer to Figure 3.6). This sequence differs from another human rp S6 cDNA at nucleotides 463 and 467 (refer to Figure 3.15).

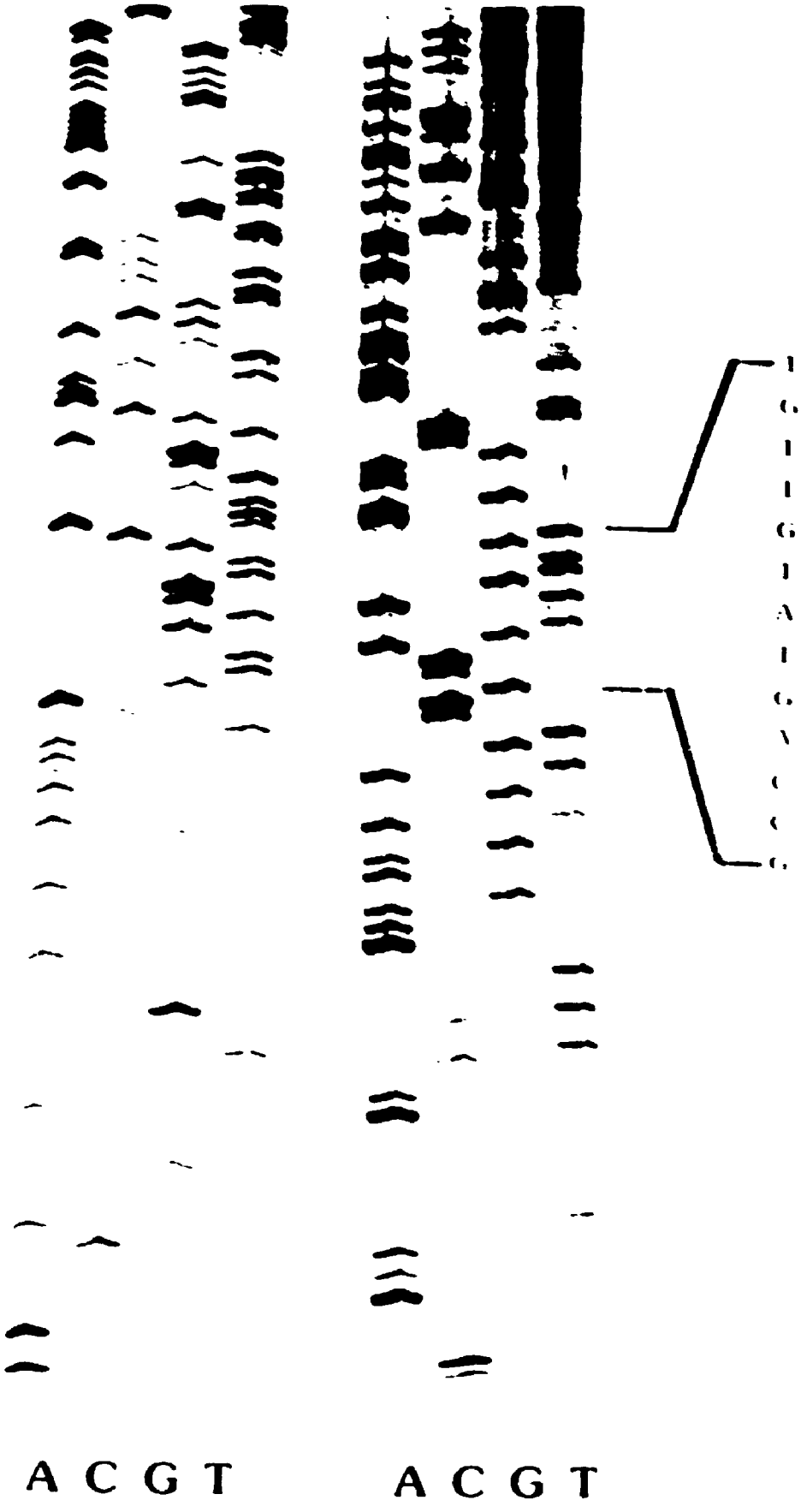

A C G T
A C G T

Figure 3.6. The cDNA sequence and predicted amino acid sequence of human ribosomal protein S6 as inferred from the cDNA inserts of pHS6-1, pHS6-2A and pHS6-2B (refer to section 3.3. Numbering is from the first residue of the initiation codon. The designated nt is aligned with the last digit of each number.

```
                                                  Met Lys Leu Asn Ile Ser Phe Pro
                                                  ATG AAG CTG AAT ATC TCT TTC CCA
                                                   1                          20

Ala Thr Gly Leu Gln Lys Leu Ile Glu Val Asp Asp Glu Arg Lys
GCT ACT GGC TTG CAG AAA CTT ATT GAA GTG GAT GAT GAA CGT AAA
                        40                  60

Leu Arg Thr Phe Tyr Glu Lys Arg Met Ala Thr Glu Val Ala Ala
CTT CGT ACT TTC TAT GAG AAG CGT ATG GCA ACA GAA GTT GCT GCT
                    80                  100

Asp Ala Leu Gly Glu Glu Trp Lys Gly Tyr Val Val Arg Ile Ser
GAC GCT CTG GGT GAA GAG TGG AAG GGT TAT GTG GTT CGA ATT AGT
        120                 140

Gly Gly Asn Asp Lys Gln Gly Phe Pro Met Lys Met Arg Gly Val Leu
GGT GGC AAC GAC AAA CAA GGT TTC CCT ATG AAG ATG CGT GGT GTT TTG
                    160                 180

Thr Ala Gly Arg Val Arg Leu Leu Leu Ser Lys Gly Asn Ser Val Asp
ACT GCT GGT CGT GTC CGT TTG TTA CTG AGT AAG GGA GAT TCC GTT GAT
                220                 240

Tyr Arg Pro Arg Arg Thr Gly Glu Arg Lys Arg Lys Ser Val Arg
TAC AGA CCA AGG AGA ACT GGA GAA AGA AAG AGA AAA TCA GTT CGT
            260                 280

Gly Cys Ile Val Asp Ala Asn Leu Ser Val Leu Asn Leu Val Ile
GGT TGC ATT GTG GAT GCA AAT CTG AGT GTT CTC AAC TTG GTT ATT
        300                 320

Val Lys Lys Gly Glu Lys Asp Ile Pro Gly Leu Thr Asp Thr Thr
GTA AAA AAA GGA GAG AAG GAT ATT CCT GGA CTG ACT GAT ACT ACA
                    340                 360

Val Pro Arg Arg Leu Gly Pro Lys Arg Ala Ser Arg Ile Arg Lys
GTG CCT CGC CGC CTG GGC CCT AAA AGA GCT AGT AGA ATC CGC AAA
                380                 400

Leu Phe Asn Leu Ser Lys Glu Asp Asp Val Arg Gln Tyr Val Val
CTT TTC AAT CTC TCT AAA GAA GAT GAT GTG CGC CAG TAT GTT GTA
            420                 440

Arg Lys Pro Leu Asn Lys Glu Gly Lys Lys Pro Arg Thr Lys Ala
AGA AAG CCT TTA AAT AAA GAA GGT AAG AAA CCT AGG ACC AAA GCA
        460                 480

Pro Lys Ile Gln Arg Leu Val Thr Pro Arg Val Leu Gln Met Lys
CCC AAG ATT CAG CGT CTT GTT ACT CCA CGT GTG CTG CAG ATG AAA
                    500                 520

Arg Arg Arg Ile Ala Leu Lys Lys Gln Arg Thr Lys Lys Asn Lys
CGG CGG CGT ATT GCT CTG AAG AAG CAG CGT ACC AAG AAA AAT AAA
                540                 560

Glu Glu Ala Ala Glu Tyr Ala Lys Leu Leu Ala Lys Arg Met Lys
GAA GAG GCT GCA GAA TAT GCT AAA CTT TTG GCC AAG AGA ATG AAG
            580                 600

Glu Ala Lys Glu Lys Arg Gln Glu Gln Ile Ala Lys Arg Arg Arg
GAG GCT AAG GAG AAG CGC CAG GAA CAA ATT GCC AAG AGA CGC AGA
        620                 640

Leu Ser Ser Leu Arg Ala Ser Thr Ser Lys Ser Glu Ser Ser Gln
CTT TCC TCT CTG CGA GCT TCT ACT TCT AAG TCT GAA TCC AGT CAG
                    660                 680

Lys
AAA TAGATTTTTTTGAGTAAGAAATAAATAAGATCAGATCTCTGA
            700                 720
```

to-10 allowing nucleotides -11 to -16 to pair with +7 to +12 (see Figure 3.7). The presence of such a secondary structure at this position may hinder the progress of reverse transcriptase along the mRNA and could explain why the inserts in pHS6-1 and pHS6-2B are shorter than full length in the libraries tested.

d. Amino Acid Sequence

The inferred amino acid sequence of human rp S6 displays regions of high conservation with yeast rp S10, separated by regions of divergence as shown in the dot matrix analysis of Figure 3.8. Overall, with minimal gapping and sliding, 63% of the amino acids are identical, while a further 18% are functionally related. Although the amino acid sequences are strikingly similar, one stretch covering 17 identical amino acids, the longest stretch of identity at the nucleotide level is only 20 nt (+166 to +185).

There are, however, two regions in which the amino acid sequences of human rp S6 and yeast rp S10 are quite divergent. The first region spans amino acids 161 to 169 in yeast rp S10 and amino acids 161 to 170 in human rp S6 (see Figure 3.9). The only matches in this stretch occur at the lysines at positions 164 and 167 in both amino acid sequences. Interestingly, the lysine-glycine-

Figure 3.7. Proposed hairpin loop structure at the 5' end of a human rp S6 coding mRNA based on the corresponding cDNA sequence. The loop is centred about a G in the -1 position and has an internal loop at the -9 position (refer to Figure 3.6). The start codon is designated by a dashed box.

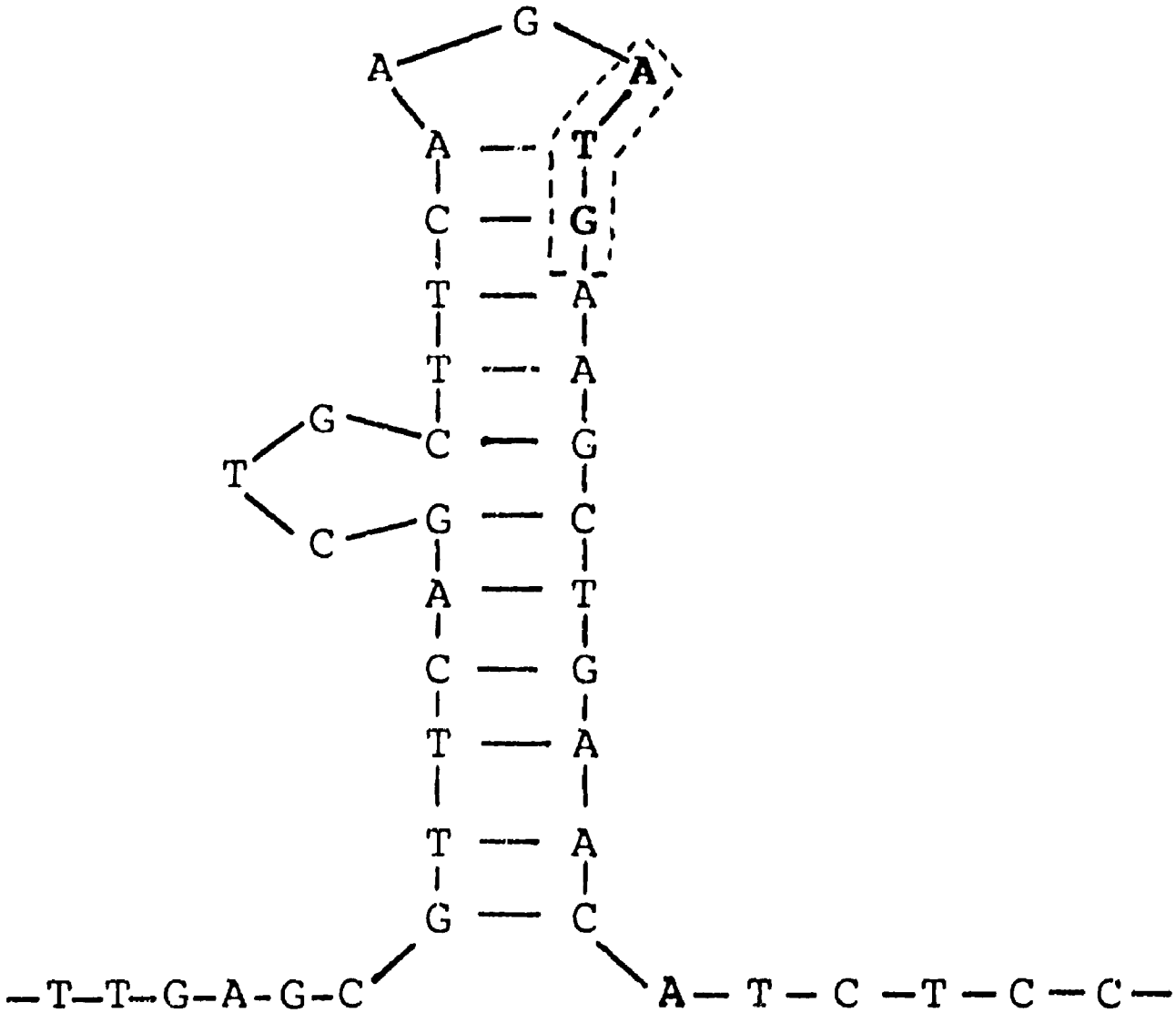
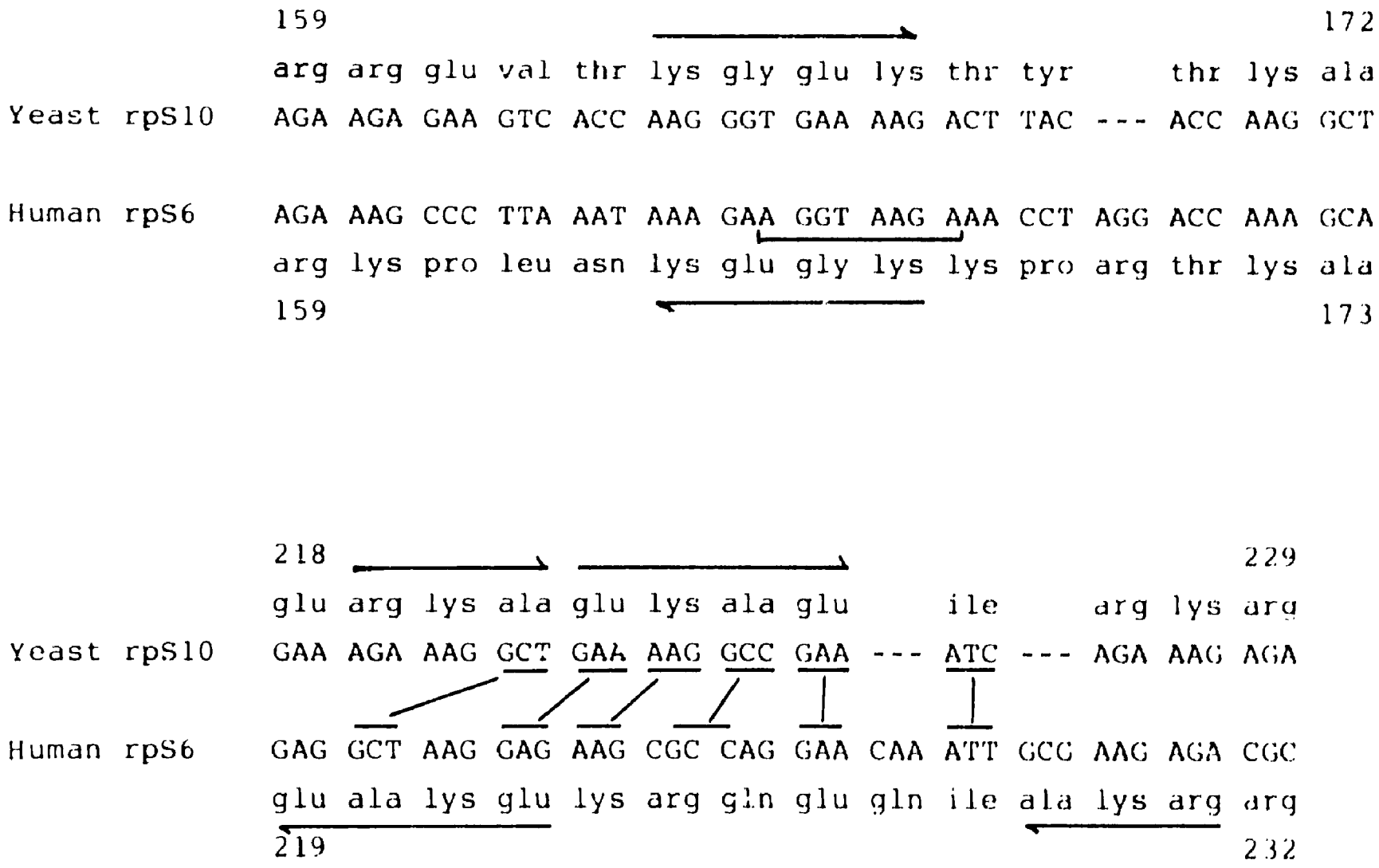

G
A
A
A
T
C
G
T
A
T
A
G
C
T
G
C
G
C
A
T
C
G
C
T
G
A
T
A
G
T
A
C
—T—T—G—A—G—C
A— T — C —T — C — C —

Figure 3.8. Comparison of amino acid homology between human rpS6 (Fig. 3.6) and yeast rpS10 (Leer et al., 1982; 1985). The two dimensional plot compares the two amino acid sequences beginning at aa residue 1 of each such that only a perfect match of 4 amino acids between each sequence is scored (Lagrimini et al., 1984).

Figure 3.10. Hydropathicity plots of yeast rp S10 and human rp S6. The plots were developed by a computer program that uses a moving-segment to determine the average hydropathy within a segment of predetermined length (n) as it advances through the sequence (in this case n=9). The consecutive scores are plotted from the amino to carboxy terminus. The scale is based on experimental observations of the hydrophobic and hydrophilic properties of the 20 amino acids. The Y axis is a function of arbitrary N values and reflects relative hydrophobicity and hydrophilicity.

Figure 3.9. Points of interest in the sequences of pHS6-2b and a yeast gene encoding rp S10 (Leer et al., 1982). Two regions of the yeast and human sequences are presented. The upper panel shows a 5' splice site consensus sequence in the human DNA sequence (Green, 1986). The lower panel shows codon homology between the yeast and human sequences that is interrupted by nucleotide insertions in the human sequence. Arrows indicate short stretches of amino acid sequence inversions between yeast and human.


159
172
arg arg glu val thr lys gly glu lys thr tyr thr lys ala
Yeast rpS10 AGA AGA GAA GTC ACC AAG GGT GAA AAG ACT TAC --- ACC AAG GCT
Human rpS6 AGA AAG CCC TTA AAT AAA GAA GGT AAG AAA CCT AGG ACC AAA GCA
arg lys pro leu asn lys glu gly lys lys pro arg thr lys ala
159
173
218
229
glu arg lys ala glu lys ala glu ile arg lys arg
Yeast rpS10 GAA AGA AAG GCT GAA AAG GCC GAA --- ATC --- AGA AAG AGA
Human rpS6 GAG GCT AAG GAG AAG CGC CAG GAA CAA ATT GCG AAG AGA CGC
glu ala lys glu lys arg gln glu gln ile ala lys arg arg
219
232

glutamate-lysine sequence found in yeast rp S10 from amino acid 164 to 167 is the reverse of the sequence found at the same positions in the human rp S6 sequence. Another intriguing feature of this region occurs in the nucleotide sequence of human rp S6. The sequence from nucleotides 495 to 502 is AGGUAAGA which, with the exception of the last adenine (at nucleotide 502), conforms to the consensus sequence of the 5' splice site for pre-mRNA processing which is AGGTAAGU (Green, 1986).

The second region of divergence spans amino acids 219 to 228 in yeast rp S10 and amino acids 220 to 231 in the human rp S6 sequence (see Figure 3.9). In this area yeast rp S10 has two fewer amino acids than the corresponding human sequence; however, the codons for the yeast rp S10 sequence are contained within the human rp S6 nucleotide sequence along with some nucleotide insertions. If, on the other hand, one considers only the amino acid sequences in this region one can find inversions of one sequence with respect to the other. The arginine-lysine-alanine sequence from amino acids 219 to 221 in yeast rp S10 is found in reverse in human rp S6 at amino acids 229 to 231 (see Figure 3.9). Similarly, the glutamate-lysine-alanine-glutamate residues at amino acids 222 to 225 in yeast rp S10 are found in reverse in human rp S6 at amino acids 219 to 222.

Whether or not these areas of divergent amino acid sequence are indicative of functional differences between yeast rp S10 and human rp S6 cannot be determined by available data. Hydropathic plots of the two amino acid sequences show that they have very similar hydrophobic and hydrophilic domains (see Figure 3.10). It is especially noteworthy that there are no significant differences between the yeast rp S10 and human rp S6 plots even in the regions of the aforementioned amino acid divergences. The only major difference is due to the additional ten C-terminal amino acids in human rp S6.

e. Size and Relative Abundance of rp S6 mRNA

Northern blot hybridization (Figure 3.11) shows that rp S6 mRNA is approximately 1 kb in length and, for comparison, human rp S11 is mRNA about 0.75 kb. Lane e shows that a yeast rp S10 probe cannot hybridize to human rp S6 mRNA even though the probe contained that portion of the yeast rp S10 gene exhibiting the longest stretch of identity with human rp S6 (20 residues).

In order to determine the relative abundance of rp S6 and rp S11 mRNAs densitometric analysis of slot-blot data was performed. Dilutions of total E. coli RNA, polyadenylated Hela S3 RNA and Hela S3 RNA minus the polyadenylated fraction were applied to a nylon membrane

Figure 3.10. Hydropathicity plots of yeast rp S10 and human rp S6. The plots were developed by a computer program that uses a moving-segment to determine the average hydropathy within a segment of predetermined length (n) as it advances through the sequence (in this case n=9). The consecutive scores are plotted from the amino to carboxy terminus. The scale is based on experimental observations of the hydrophobic and hydrophilic properties of the 20 amino acids. The Y axis is a function of arbitrary N values and reflects relative hydrophobicity and hydrophilicity.

Hydropathicity Plot

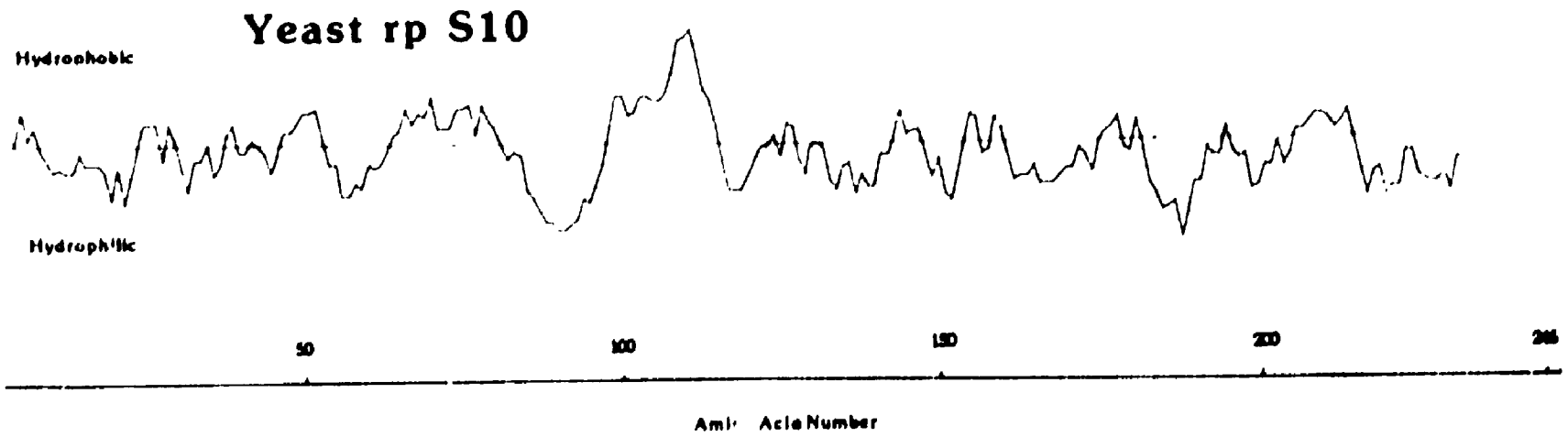
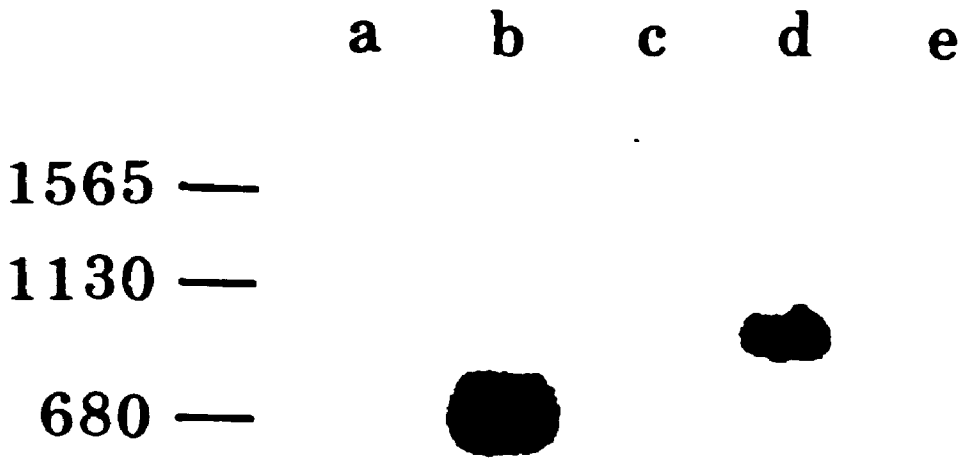


Figure 3.11. Size determination of human rpS6 and rpS11 mRNAs. Samples (5 ug each) of polyadenylated (lanes B, D and E) or non-polyadenylated (lanes A and C) Hela cell RNA were subjected to "Northern" blotting as described in section 3.3. Lanes A and B were probed with a cDNA corresponding to human rpS11 (Lott and Mackie, 1988). lanes C and D were probed with a cDNA corresponding to nt residues 213 to 401 of pHS6-1 while lane E was probed with a fragment of yeast rpS10 DNA. Numbers in the left margin give the sizes in nt of denatured DNA fragments of known sequence which were detected by staining of the gel prior to transfer.


a
b
c
d
e
1565
1130
680

Southern hybridizations of other ribosomal protein cDNA and implies the existence of multiple copies of the rp S6-coding gene(s)/pseudogene(s).

Figure 3.12. Autoradiograph of a slot blot hybridization. Two panels of slots were prepared such that each contained E. coli RNA, human polyadenylated RNA and human RNA minus the polyadenylated fraction. The RNA samples were applied to the filter in quantities of 2.5 ug, 0.625 ug and 0.156 ug. A lane containing no RNA was also present. One panel was probed with a nick-translated fragment of pRS-11 and the other was probed with a similarly prepared fragment of pHS6-1.

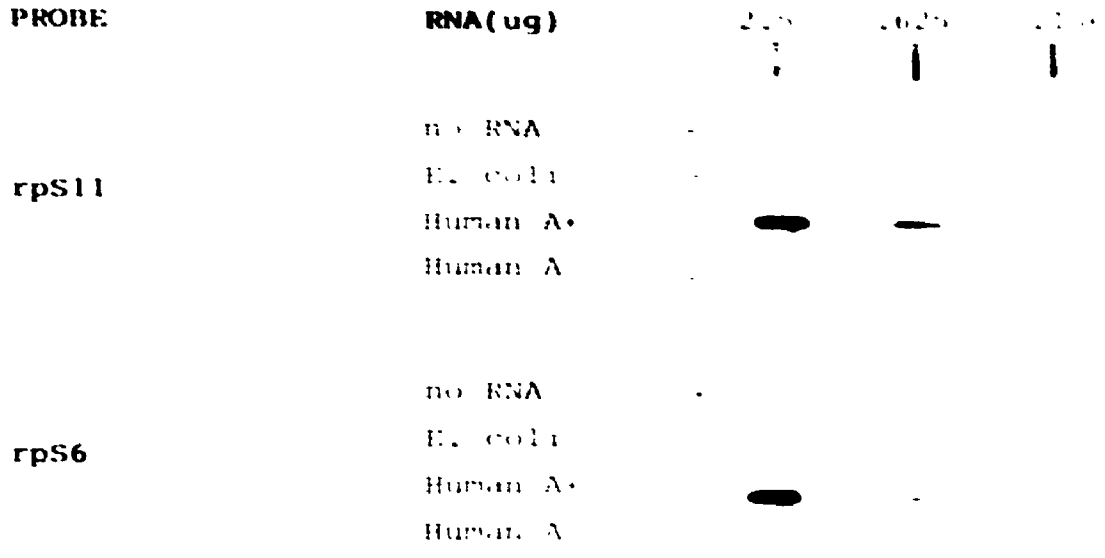

PROBE
RNA(ug)
rpS11
rpS6

Figure 3.13. Densitometry of Northern blot hybridizations. Two different Northern blot hybridizations were scanned with a densitometer to determine the relative abundance of human rpS11 and human rp s6 mRNA in the polyadenylated fraction. The peaks were produced by positions in the autoradiograph that correspond to either the rp S11 mRNA or rp S6 mRNA.

	Probe	Peak Height	Base* Line	Net Peak Height	Area** Under Peak	Ratio of Area(S11) / Area(S6)
NORTHERN #1	rpS11	1.28	.39	.89	5.53	2.3
	rpS6	.87	.36	.51	2.39	
NORTHERN #2	rpS11	1.48	.40	1.08	4.85	2.1
	rpS6	.94	.37	.57	2.36	

* The base line is the average of the 16 lowest data points.

** Area under Peak is the product of the maximum absorbance minus the minimum absorbance and the width of the peak in millimetres.

suggest that the absolute levels of different rp mRNAs may differ even if the synthesis of their protein products is coordinately controlled.

f. Reiteration of the rp S6-coding Gene

Others (Dudov and Perry, 1984; Wiedemann and Perry, 1984; Nakamichi et al., 1986; Monk et al., 1981; Kuwano et al., 1985; Peled-Yalif et al., 1984) have reported that mammalian ribosomal proteins are represented by at least one functional gene and by multiple pseudogenes, many of which represent fully processed transcripts. Human genomic DNA was digested with restriction enzymes that do not recognize sequences within the cDNA (Eco RI, Hind III and Bgl II) and with Pst I which has two sites within the cDNA. The resultant blot was hybridized with nick-translated inserts from pHS6-2A and pHS6-2B and then exposed to film. The autoradiogram in Figure 3.14 shows that each lane contains at least seven bands and that the Pst I lane (lane 1) has at least ten. The intensities vary between bands in the same lane in proportion to the amount of labelled probe that annealed to each DNA fragment. Faint bands may indicate DNA fragmets which contain relatively short stretches of coding sequence bounded by introns containing site(s) for one of the enzymes tested. This result is consistent with the

Figure 3.14 (A). Analysis of ribosomal protein S6 genes in the human genome. Each lane contains 15 ug of human genomic DNA that was digested with the restriction enzyme indicated below. The DNA was subjected to "Southern" blotting as described in section d of Materials and Methods, probed with 32P-nick-translated cDNA inserts from pHS6-2A and pHS6-2B, and visualized by autoradiography. The endonucleases are: P, PstI; E, EcoRI; H, HindIII; B, BglII. Markers in the right hand margin give sizes of a HindIII digest of λDNA [illegible] kb detected by staining of gel prior to transfer.

Three exposures were made of the autoradiograph in order to detect all bands. The photograph represents a compromise between detection of faint bands and over exposure of dark bands.

(B). A diagram to clarify the position of each band in the autoradiograph in (A) is presented. Each dash relates to a band in (A) such that super imposition of (B) onto (A) would align the dashes with the corresponding hybridization signal.

(A)

P E H B

- 23.1
- 9.4
- 6.7
- 4.4
- 2.3
- 2.0
- 0.6

(B)

P E H B

- 23.1
- 9.4
- 6.7
- 4.4
- 2.3
- 2.0
- 0.6

Southern hybridizations of other ribosomal protein cDNA and implies the existence of multiple copies of the rp S6-coding gene(s)/pseudogene(s).

3.4 Discussion

a. Introduction

A mixed oligodeoxynucleotide probe was employed to isolate a human cloned cDNA whose predicted amino acid sequence possesses remarkable homology to yeast rp S10. Furthermore, this cDNA was employed to rescreen the library from which it originated and to screen other cDNA libraries to yield two more related cloned cDNAs. The amino acid sequence that these clones predict is identical to the published mouse rp S6 sequence (Lalanne et al., 1987). We conclude that together these three cloned cDNAs represent the entire coding sequence of human rp S6. The inability to isolate a human rp S6 coding cDNA or genomic clone using the approaches described in chapter 2 was due to the low nucleotide sequence homology between the coding sequences of yeast rp S10 (Leer et al., 1982; 1985) and human rp S6 (Lott and Mackie, 1988). Although there are also areas of amino acid divergence between the two proteins, yeast rp S10 and human rp S6 appear to be functionally related (Zinker and Warner, 1976) suggesting that there may be islands of conserved amino acid sequence, especially in any domains important for the function of the protein. This is supported by the hydropathicity plots (see Figure 3.10).

Within conserved sequences there may be stretches of identical amino acids sufficiently long that one could confidently design oligodeoxynucleotide probes with a high probability of identifying the desired clones in a cDNA or genomic library.

It is noteworthy that the isolation of cDNA or genomic sequences employing an oligodeoxynucleotide probe as short as a 14 mer and possessing 32-fold degeneracy has not been reported elsewhere. The high specificity of this short probe is also quite remarkable.

The striking amino acid sequence homology between yeast rp S10 (Leer et al., 1982) and human rp S6 (Figure 3.6) and the nearly perfect N and C terminal amino acid sequence homology between rat rp S6 (Wool, 1979; Wettenhall and Morgan, 1984) and human rp S6 led us to believe that the cloned cDNAs contained in pHS6-1, pHS6-2a and pHS6-2b encoded human rp S6. Recently, this conclusion has been supported by the published cDNA and predicted amino acid sequences of mouse rp S6 (Lalanne et al., 1987) and rat rp S6 (Chan and Wool, 1988) which both share 100% amino acid sequence homology with human rp S6. Identification by comparison of protein sequence data with the amino acid sequence predicted by a cloned cDNA will not always yield 100% homology. For example, the sequence of rat rp S6 as determined by protein sequencing (Wettenhall et al., 1988) and the amino acid sequence

predicted by a cloned cDNA encoding rat rp S6 (Chan and Wool, 1988) differ at seven positions.

Another method for demonstrating the authenticity of the cloned cDNAs would compare the in vitro transcription and translation protein product of the cloned cDNA with native human rp S6 employing enzyme digestion maps and two dimensional electrophoresis. This was not performed as there was no straightforward means of joining portions of pHS6-2a and pHS6-2b to yield a single clone encompassing the entire coding region. The study of ribosomal protein gene regulation will require the isolation of the expressed intron-containing gene encoding human rp S6. Restriction enzyme sites common to both the expressed gene and to pHS6-2b may permit a more favourable cDNA construction which would include the entire 5' untranslated region.

b. Nucleotide Sequence

The difficulty in isolating full length cDNA clones encoding human ribosomal protein S6 may be explained partly by the sequence of the cDNA insert in pHS6-2A. The inserts found in pHS6-1 and pHS6-2B were short of full length and neither covered the region of the proposed hairpin loop (refer to Figure 3.7) while the insert in pHS6-2A which did carry sequence 5' to the hairpin loop

also contained an inversion and deletion. We believe that this clone may have arisen from an enzymatic error during synthesis of the first strand of the cDNA caused by hairpin formation at the 5' end of the mRNA and consequently, the creation of an inverted repeat after synthesis of the second strand. Thereafter, the inverted repeat may have been deleted by the bacterial host during propagation of the phage. It is interesting to note that neither the mouse (Lalanne et al., 1987) nor rat rp S6 (Chan et al., 1988) encoding sequences can form stable hairpin loops in the same region as their human counterpart. This may have interesting regulatory or translational implications which can be investigated in the future.

c. Size and Relative Abundance of rp S6 mRNA

Restriction maps and nucleotide sequence data from pHS6-1 and pHS6-2B imply that the cloned polyadenylated tails in these cDNAs are 125 bp and 135 bp respectively. Taking into account the data from sequencing and the Northern transfer hybridization, the human sequence would predict a 5' untranslated region similar in size to those of other published mammalian rp mRNAs (Dudov and Perry, 1984; Wiedemann and Perry, 1984; Rhoads et al., 1986), that is about 50 to 70 residues.

The densitometric analysis of the Northern blot hybridizations reveal that rp S11 mRNA is approximately 2.2 times as abundant as rp S6 mRNA. This argues against Roufa's idea that most rp mRNAs are equally abundant (D. J. Roufa, pers. commun. to G. A. Mackie).

The difference in the amount of the two ribosomal protein messages may indicate different types of post-transcriptional regulation of human ribosomal protein synthesis such as differences in translation efficiency or mRNA stability. In yeast (Warner et al., 1985), the levels of individual ribosomal proteins may be regulated by protein turnover, inhibition of splicing of the primary transcript, distribution between the polyribosomal and non-polyribosomal fractions, or some as yet unrecognized method of control.

d. Comparative Examination of Mammalian Ribosomal Protein S6 cDNA Sequences

Recently, the sequences of three other mammalian ribosomal protein S6 cDNAs have been reported. These include the ribosomal protein S6 cDNA sequences of mouse (Lalanne et al., 1987), rat (Chan and Wool, 1988) and another human cloned cDNA (Heinze et al., 1988) which were published during the preparation for publication of

our human ribosomal protein S6 cDNA sequence (Lott and Mackie, 1988) or during the preparation of this thesis.

The cloned cDNAs for mouse, rat and our human clone predict the same amino acid sequence. With respect to the coding regions, the mouse nucleotide sequence differs from that of our human sequence in the first position in 6 codons and in the third position in 85 codons. The published rat cDNA sequence differs from our human sequence in the first position in 5 codons and in the third position in 79 codons.

The two human ribosomal protein S6 cDNA sequences differ in nuclectide sequence in 6 positions within the coding region leading to 4 amino acid differences (Figure 3.15). In the 3' untranslated region the two cDNAs are mismatched at one nucleotide and our cDNA has one fewer nucleotide. The differences between this rp S6 sequence and that of Heinze et al. (1988) could reflect polymorphisms. The rp S6 sequences from mouse (Lalanne et al, 1987), rat (Chan and Wool, 1988) and our human clones (Lott and Mackie, 1988) were, however, derived from mouse T cells, rat liver, and human fibroblasts or human placenta respectively and all encode the same amino acid sequence. The observed species and tissue conservation of the rp S6 amino acid sequence and the lack of evidence for tissue-specific ribosomal proteins strongly suggests that the sequence provided by Heinze et al. (1988) may

Figure 3.15. Comparison of selected codons and predicted amino acids of mammalian rp S6 sequences. The rp S6 sequences shown are mouse (Lalanne et al., 1987), rat (Chan and Wool, 1988), human(a) (Lott and Mackie, 1988) and human(b) (Heinze et al., 1988). The table shows the predicted amino acid differences among the four mammalian sequences and also nucleotide differences between the two human rp S6 cDNAs.

AMINO ACID	CODON AND PREDICTED AMINO ACID			
POSITION	MOUSE	RAT	HUMAN(a)	HUMAN(b)
135	P· CCT	Pro CCT	Pro CCC	Pro CCA
145	Phe TTT	Phe TTT	Phe TTC	Phe TTT
156	Tyr TAT	Tyr TAT	Tyr TAT	Cys TGT
155	Gln CAG	Gln CAG	Gln CAG	Glu GAG
168	Lys AAG	Lys AAG	Lys AAA	Arg AGA
219	Glu GAA	Glu GAA	Glu GAG	Gln CAG

contain errors. Recently another group has isolated a human rp S6 cloned cDNA whose nucleotide sequence matches ours in all positions (A. Metspalu, pers. commun. to G. A. Mackie).

Heinze et al. (1988) did not report any of the sequence of the 5' untranslated region and, therefore, comparisons in this area are not possible. The lack of 5' sequence is of interest since the cDNA reported by Metspalu et al. (pers. commun. to G. A. Mackie) also lacks the adenine of the start codon and does not contain any of the 5' untranslated region as well. The inability of others to obtain human clones that extend beyond the start codon in a 5' direction lends support to the proposal that a hairpin loop in the area of the start codon may inhibit the progress of reverse transcriptase during first strand synthesis.

e. Speculations on Intron-Exon Junctions in the Human rp S6 cDNA Sequence

One cannot confidently determine the points of splicing in mature mRNAs without having the sequences of both the functional gene and the cloned cDNA. In this work only the sequence of the cloned cDNA encoding human rp S6 was determined. A comparison of this sequence with that of yeast rp S10 (Leer et al., 1982; 1985) reveals

some interesting features in the regions of greatest amino acid divergence. The presence of a 5'splice consensus sequence in the stretch of divergent amino acids from 161 to 170 in the human sequence (Figure 3.8) may indicate that during evolution amino acid diversity may have resulted in changes in the splice site. In the second region of amino acid divergence, from amino acids 220 to 231 in the human sequence (Figure 3.9), the corresponding nucleotide sequence has six more bases than the comparable yeast sequence and yet the yeast rp S10 codons can be found within the human sequence. It is tempting to speculate that this region contains the union of two exons and that replication errors have allowed the accumulation of six more nucleotides than are present in yeast.

Considerations such as these can be useful in the identification of the functional intron-containing gene. The results of the southern blot (Figure 3.13) demonstrate that there are several rp S6 coding sequences within the human genome. In order to determine which of these contains introns, restriction enzymes that have recognition sequences contained within the rp S6 cDNA and that flank a possible intron-exon junction, could serve to locate the intron-containg gene in a genomic library by restriction fragment length changes. An adjunct procedure would be to test all isolates of a genomic

library that contain rp S6 coding sequences for CG islands. These islands which are clusters of relatively short sequences containing high percentages of C and G residues appear to flank some functional eukaryotic genes. Some functional genes have been identified by using a battery of CG cutting enzymes to determine the locations of CG islands then searching the region between for open reading frames (Lindsay and Bird, 1987; Page et al., 1987).

The functional gene encoding human ribosomal protein S6 remains to be isolated but I have isolated three cDNA clones that together contain the entire coding sequence of human rp S6, all of the 3' untranslated region and 27 nt of the 5' leader. These clones will be valuable in helping to locate the functional rp S6 gene(s) and later in the study of gene expression and regulation.

REFERENCES

BAILEY, J. M. and DAVIDSON, N. (1976). Methlymercury as a reversible denaturing agent for agarose gel eletrophoresis. Anal. Biochem. 70, 75-82.

BALLAL, N. R., KANG, Y-J., OLSON, M. J. and BUSCH, H. (1975). Changes in nucleolar proteins and their phosphorylation patterns during liver regeneration. J. Biol. Chem. 250, 5921-5925.

BEAUCAGE, S. L. and CARUTHERS, M. H. (1981). Deoxynucleoside phosphoramidites. A new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. 22, 1859-1862.

BENTON, W. D. and DAVIES, R. W. (1977). Screening lambda gt recombinant clones by hybridization to single plaques in situ. Science 196, 180-182.

BIGGIN, M.D., GIBSON, J.T. and HONG, G.F. (1983). Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. Proc. Nat. Acad. Sci. USA 80, 3963-3965.

BIRD, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucl. Acids Res. 8, 81499-1504.

BLENIS, J. and ERIKSON, R.L. (1984). Phosphorylation of the ribosomal protein S6 is elevated in cells transformed by a variety of tumor viruses. J. Virol. 50, 966-969.

BLENIS, J. and ERIKSON, R. L. (1985). Regulation of ribosomal protein S6 kinase activity by the rous sarcoma virus transforming protein, serum, or phorbol ester. Proc. Natl. Acad. Sci. USA 82, 7621-7625.

BLENIS, J., SPIVACK, J. G., and ERIKSON, R. L. (1984). Phorbol ester, serum and Rous sarcoma virus transforming gene product induce similar phosphorylations of ribosomal protein S6. Proc. Nat. Acad. Sci. USA 81, 6408-6412.

BLENIS, J., SPIVAK, J. G. and ERIKSON, R. L. (1984). Phorbol ester, serum and Rous sarcoma virus transforming gene product induce similar phosphorylations of ribosomal protein S6. Proc. Natl. Acad. Sci. USA 81, 6408-6412.

BOLIVAR, F. (1978). Construction and characterization of new cloning vehicles. III. derivatives of plasmid pBR322 carrying unique Eco RI sites for selection of Eco RI generated recombinant DNA molecules. Gene 4, 121-136.

BOZZINI, I., BECCARI, E., LUO, Z. X., AMALDI, F., PIERANDREI-AMALDI, P. and CAMPIONI, N. (1981). Xenopus laevis ribosomal protein genes: isolation of recombinant cDNA clones and study of the genomic organization. Nucl. Acids Res. 9, 1069-1086.

BOZZONI, I., FRAGAPANE, P., ANNESI, F., PIERANDREI-AMALDI, P., AMALDI, F. and BECCARI, E. (1984). Expression of two Xenopus laevis ribosomal protein genes in injected frog oocytes. J. Mol. Biol. 180, 987-1005.

BURKHARD, S. J. and TRAUGH, J. A. (1983). Changes in ribosome function by cAMP-dependent and cAMP-independent phosphorylation of ribosomal protein S6. J. Biol. Chem. 258, 14003-14008.

BUTTEREIT, D., PFLUGFELDER, G. and GRUMMT, I. (1985). Growth dependent regulation of rRNA synthesis is mediated by a transcription iniation factor (TIF-IA). Nucleic Acids Res. 13, 8165-8180.

CAFFARELLI, E., FRAGPANE, P., GEHRING, C. and BOZZONI, I. (1987). The accumulation of mature RNA for the Xenopus laevis ribosomal protein L1 is controlled at the level of splicing and turnover of the precursor RNA. Embo J. 6, 3493-3498.

CANAANI, E. and AARONSON, A. (1979). Restriction enzyme analysis of mouse cellular type C viral DNA: Emergence of new viral sequences in spontaneous AKR/J lymphomas. Proc. Nat. Acad. Sci. USA 76, 1677-1681.

CASEY, J. and DAVIDSON, N. (1977). Rates of formation and thermal stability of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucl. Acids Res. 4, 1539-1552.

CHAN, Y-L., LIN, A., PAZ, V. and WOOL, I. (1987). The primary structure of rat ribsomal protein S8. Nucleic Acids Res. 15, 9451-9459.

CHAN, Y-L. and WOOL, I. (1988). The primary structure of rat ribosomal protein S6. J. Biol. Chem. 263, 2891-2896.

CHEN, I., DIXIT, A., RHOADS, D., and ROUFA D. J. (1986). Homologous ribosomal proteins in bacteria, yeast, and humans. Proc. Nat. Acad. Sci. USA 83, 6907-6911.

CHOOI, W. Y. and LEIBY, K. R. (1981). An electron microscopic method for localization of ribosomal proteins during transcription of ribosomal DNA: A method for studying protein assembly. Proc. Natl. Acad. Sci. USA 78, 4823-4827.

CILIBERTO, G., RAUGEI, G., COSTANZO, F., DNETE, L. and CORTESE, R. (1983). Common and interchangeable elements in the promoters of genes transcribed by RNA polymersae III. Cell 32, 726-733.

DABEVA, M. D. POST-BEITTENMILLER, M. A. and WARNER, J. R. (1986). Autogenous regulation of splicing of the transcript of a yeast ribosomal protein gene. Proc. Natl. Acad. Sci. USA 83, 5854-5857.

DAVIES, M. S., HENNEY, A., WARD, W. and CRAIG, R. K. (1986). Characterization of an mRNA encoding a human ribosomal protein homologous to the yeast L44 ribosomal protein. Gene 45, 183-191.

DEININGER, P., JOLLY, D., RUBIN, C., FRIEDMANN, T. and SCHMID, C. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J. Mol. Biol. 151, 17-33.

DELAUNAY, J. and SCHAPIRA, G. (1974). Immunogenic properties of eukaryotic ribosomal proteins. Biochem and Biophys Acta 386, 138-141.

DENHARDT, D. T. (1966). A membrane filter technique for the detection of complimentary DNA. Biochem. Biophys. Res. Commun. 23, 641-646.

DERYNCK, R., ROBERTS, A. B., WINKLER, M. E., CHEN, E. Y. and GOEDDEL, D. V. (1984). Human transforming growth factor-alpha: Precursor structure and expression in E. coli. Cell 38, 287-297.

D'EUSTACHIO, D., MEYUHAS, O., RUDDLE, F. and PERRY, R. P. (1981). Chromosomal distribution of ribosomal protein genes in the mouse. Cell 24, 307-312.

DONOVAN, D. M. and PEARSON, N. J. (1986). Transcriptional regulation of ribosomal proteins during a nutritional upshift in Saccharomyces cerevisiae. Mol. and Cell. Biol. 6, 2429-2435.

DUDOV, K. P. and PERRY, R. P. (1984). The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. Cell 37, 457-468.

DYNAN, W. S. and TIJAN, R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell 32, 669-680.

EIL, C. and WOOL, I. G., (1973). Characteristics of the protein kinase reaction and studies of the structure of phosphorylated ribosomes. J. Biol. Chem. 248, 5122 5129.

ERICKSON, J. M., RUSHFORD, C. L., DORNEY, D. J., WILSON, G. N. and SCHMICKEL, R. D. (1981). Structure and variation of human ribsomal DNA: molecular analysis of cloned fragments. Gene 16, 1-9.

EVANS, S. W. and FARRAR, W. L., (1987). Interleukin 2 and diacylglycerol stimulate phosphorylation of 40 S ribosomal protein S6 protein. J. Biol. Chem. 262, 4624-4630.

FARRAR, W. L. and ANDERSON, W. B. (1985). Interleukin-2 stimulates association of protein kinase C with plasma membrane. Nature 315, 233-235.

FEINBERG, A. P. and VOGELSTEIN, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem. 132, 6-13.

FIRE, A., SAMUELS, M. and SHARP, P. A. (1984). Interactions between RNA polymerase II, factors, and template leading to accurate transription. J. Biol Chem. 259, 2509-2516.

FISCHER, N., STOFFLER, G., and WOOL, I. G. (1978). Immunological comparison of the proteins of chicken and rat liver ribosomes. J. Biol. Chem. 253, 7355-7360.

FORGET, B. G. and WEISSMAN, S. M. (1967). Nucleotide sequence of KB cell 5S RNA. Science 158, 1695-1699.

FREIDMAN, D. I., OLSON, E. R., GEORGOPOULOS, C., TILLY, K., HERSKOWITZ, I. and BANUETT, F. (1984). Interactions of bacteriophage and host macromolecules in growth of bacteriophage lambda. Microbiol. Rev. 48, 299-325.

GEISSER M., TISCHENDORF, G. W., STOFFLER, G. and WITTMANN, H. G. (1973a). Immunological and electrophoretic comparison of ribosomal proteins from eight species belonging to Enterobacteriaceae. Molec. Gen. Genet. 127, 111-128.

GEISSER, M., TISHENDORF, G. W., STOFFLER, G. and WITTMAN, H. G. (1973b). Comparative immunological and electrophoretic studies on ribosomal proteins of bacilliaceae. Molec. Gen. Genet. 127, 129-145.

GONZALEZ, I.L., GORSKI, J.L., CAMPEN, T.J., DORNEY, D.J., ERICKSON, J. M., SYLVESTER, J. E. and SCHMICKEL, R. D. (1985). Variation among human 28S ribosomal RNA genes. Proc. Natl. Acad. Sci. 82, 7666-7670.

GONZALEZ, I. L. and SCHMICKEL, R. D. (1986). The human 18S ribosomal RNA gene: Evolution and stability. Am. J. Hum. Genet. 38, 419-427.

GRANTHAM, R., GAUTIER, C., GOUY, M , JACOBZONE, M. and MERCIER, R. (1981). Codon catalog usage is a genome strategy modulated for gene expression. Nucl. Acids Res. 9, 43-73.

GREEN, M. R. (1986). Pre-mRNA splicing. Ann. Rev. Genet. 20, 671-708.

GRESSNER, A. M. and WOOL, I. (1976). Influence of glucagon and cyclic adenosine 3':5'-monophosphate on the phosphorylation of rat liver ribosomal protein S6. J. Biol. Chem. 251, 1500-1504.

GRUMMT, I. and GRUMMT, F. (1974). Studies on the phosphorylation of nuclear proteins. Febs Letters . 39, 129-132.

GRUMMT, I., ROTH, E. and PAULE, M. R. (1982). Ribosomal RNA transcription in vitro is species specific. Nature 296, 173-174.

GUPTA, R. S., and SIMINOVITCH, L. (1977). The molecular basis of emetine resistance in chinese hamster ovary cells: alteration in the 40S ribsomal subunit. Cell 10, 61-66.

HAWLEY, D. K. and ROEDER, R. G. (1985). Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. J. Biol. Chem. 260, 8163-8172.

HAWLEY, D. K. and ROEDER, R. G. (1987). Functional steps in transcription initiation and reinitiation from the major late promoter in a HeLa nuclear extract. J. Biol. Chem. 262, 3452-3461.

HEFLMAN, D. M., FERAMISCO J. R., FIDDES, J. C. THOMAS, G. P. and HUGHES S. H. (1983). Indentification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. Proc. Natl. Acad. Sci. USA 80, 31-35.

HEINZE, H. ARNOLD, H. H., FISCHER, D. and KRUPPA, J. (1988). The primary structure of the human ribosomal protein S6 derived from cloned cDNA. J. Biol. Chem. 263, 4139-4144.

HOEFFLER, W. K. and ROEDER, R. G. (1985). Enhancement of RNA polymerase III transcription by the E1A gene product of adenovirus. Cell 41, 955-963.

HUET, J., COTTRELL, P., COOL, M., VIGNAIS, M.-L., THIELE, D., MARDK, C., BUHLER, J.-M., SENTENAC, A. and FROMAGEOT, P. (1985). A general upstream binding factor for genes of the yeast translation apparatus. Embo J. 4, 3539-3547.

HUGLE, B., SCHEER, U. and FRANKE, W. W. (1985). Ribocharin: A nuclear Mr 40,000 protein specific to precursor particles of the large ribosomal subunit. Cell 41, 615-627.

HWU, H. R., ROBERTS, J. W. DAVIDSON, E. H. and BRITTEN, R. J. (1986). Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc. Natl. Acad. Sci. USA 83, 3875-3879.

INGLES, C. J., BIGGS, J., WONG, J.K.-C., WEEKS, J.R. and GREENLEAF, A. L. (1983). Identification of a structural gene for a RNA polymerase II polypeptide in Drosophila melanogaster and mammalian species. Proc. Natl. Acad. Sci. USA 80, 3396-3400.

ITAKURA, K., ROSSI, J. J. and WALLACE, R. B. (1984). Synthesis and use of synthetic oligonucleotides. Ann. Rev. Biochem. 53, 323-356.

ITOH, T. H., OTAKA, E. and OSAWA, S. (1980) in Genetics and Evolution of RNA Polymerase tRNA and Ribosomes (Osawa, S., Ozeki, H., Uchida, H. and Yura, T., eds) 609-624. University of Tokyo Press, Tokyo.

ITOH, T. and WITTMANN-LIEBOLD, B. (1978). The primary structrure of protein 44 from the large subunit of yeast ribosomes. Febs Letters 96, 399-402.

JAYE, M., DE LA SALLE, H., SCHAMBER, F., BALLAND, A., KOHLI, V., FINDELI, A., TOLSTOSHEV, P. and LECOCQ, J.-P. (1983). Isolation of a human anti-haemophilic factor IX cDNA clone using a unique 52-base synthetic oligonucleotide probe deduced from the amino acid sequence of bovine factor IX. Nucleic Acids Res. 11, 2325-2335.

JOHNSON, S. P. and WARNER, J. R. (1987). Phosphorylation of the Saccharomyces cerevisiae equivalent of ribosomal protein S6 has no detectable effect on growth. Mol. Cell. Biol. 7, 1338-1345.

JOLLY, D. J., OKAYAMA, H., BERG, P., ESTY, A. C., FILPULA, D., BOHLEN, P., JOHNSON, G. G., SHIVELY, J. E., HUNKAPILLAR, T. and FRIEDMANN, T. (1983). Isolation and characterization of a full-length expressible cDNA for human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. 80, 477-483.

KABAT, D. (1971). Phosphorylation of ribosomal proteins in rabbit reticulocytes. A cell-free system with ribosomal protein kinase activity. Biochemistry 10, 197-203.

KHANDJIAN, E. W. (1987). Optimized hybridization of DNA blotted and fixed to nitrocellulose and nylon membranes. Biotechnology 5, 165-167.

KISILEVSKY, R., TREOLAR, M. A. and WEILER, L. (1984). Ribosome conformational changes associated with protein S6 phosphorylation. J. Biol. Chem. 259, 1351-1356.

KLEIN, A. and MEYUHAS, O. (1984). A multigene family of intron lacking and containing genes, encoding for mouse ribosomal protein L7. Nucl. Acids Res. 12, 3763-3775.

KOZAK, M. (1987). Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. Mol. and Cell. Biol. 7, 3438-3445.

KRUSE, C., JOHNSON, S. P. and WARNER, J. R. (1985). Phosphorylation of the yeast equivalent of ribosomal protein S6 is not essential for growth. Proc. Natl. Acad. Sci. USA 82, 7515-7519.

KUWANO, Y., NAKANISHI, O., NABESHIMA, Y., TANAKA, T. and OGATA, K. (1985) Molecular cloning and nucleotide sequence of DNA complimentary to rat ribosomal protein S26 messenger RNA. J. Biochem. 97, 983-992.

KYTE, J. and DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

LAGRINIMI, L. M., BRENTANO, S. T. and DONELSON, J. E. (1984). A DNA sequence analysis package for the IBM personal computer. Nucl. Acids Res. 12, 605-614.

LALANNE, J., LUCERO, M. and LE MOULLEC, J.: (1987). Complete sequence of mouse S6 ribosomal protein. Nucl. Acids Res. 15, 4990.

LARKIN, J. C. and WOOLFORD, J. L. (1983). Molecular cloning and analysis of the CRYI gene: a yeast ribosomal protein gene. Nucl. Acids Res. 11, 403-420.

LASSAR, A. B., HAMER, D. H. and ROEDER, R. G. (1985). Stable transcription complex on a class III gene in a minichromosome. Mol. Cell. Biol. 5, 40-45.

LASSAR, A. B., MARTIN, P.L. and ROEDER, R. G. (1983). Transcription of class III genes: formation of preinitiation complexes. Science 23, 740-748.

LASTICK, S. M. (1980). The assembly of ribosomes in HeLa cell nucleoli. Eur. J. Biochem 113, 175-182.

LAWN, R. M., FRITSCH, E. F., PARKER, R. C., BLAKE, G. and MANIATIS, T. (1978). The isolation and characterization of linked alpha-and beta-globin gene from a cloned library of human DNA. Cell 15, 1157-1174.

LEARNED, R. M., CORDES, S. and TJIAN, R. (1985). Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I. Mol. Cell. Biol. 5, 1358-1369.

LEER, R. J., VAN RAAMSDONK-DUIN, M. M. C., MOLENAAR, C. M. TH., COHEN, L. H., MAGER, W. H. and PLANTA, R. J.

(1982). The structure of the gene coding for the phophorylated ribosomal protein S10 in yeast. Nucl. Acids Res. 10, 5869-5878.

LEER, R. J., VAN RAAMSDONK-DUIN, M. M. C., MOLENAAR, C. M. TH., WITSENBOER, H. M. A., MAGER, W. H. and PLANTA, R. J. (1985) Yeast contains two functional genes coding for ribosomal protein S10. Nucl. Acids Res. 13, 5027- 5039.

LE PEUCH, C. J., BALLESTER, R. and ROSEN, O. M. (1983). Purified rat brain calcium- and phospholipid-dependent protein kinase phosphorylates ribosomal protein S6. Proc. Natl.Acad. Sci. USA 80, 6858-6862.

LIN, A., McNALLY, J. and WOOL, I. G. (1983). The primary structure of rat liver ribosomal protein L37. J. Biol. Chem. 258, 10664-10671.

LINDAHL, L. and ZENGEL, J. M. (1986). Ribosomal genes in Escherichia coli. Ann. Rev. Genet. 20, 297-326.

LINDSAY, S. and BIRD, A. P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 327, 336-338.

LISCHWE, L. M., O'LEARY, P. and BUSCH, H. (1983). Localization of nuclear phosphoproteins B23 and C23 during mitosis. Experimental Cell Res. 146, 139-149.

LOTT, J. B. and MACKIE, G. A. (1988a). Sequence of a cloned cDNA encoding human ribosomal protein S11. Nucl. Acids Res. 16, 1205.

LOTT, J. B. and MACKIE, G. A. (1988b). Isolation and characterization of cloned cDNAs that code for human ribosomal protein S6. Gene 65, 31-39.

MACKIE, G. A. (1986). Structure of the DNA distal to the gene for ribosomal protein S20 in Escherichia coli K-12: presence of a strong terminator and IS1 element. Nucl. Acids Res. 14, 6965-6981.

MALLER, J. L., PIKE, L. J., FREIDENBERG, G. R., COPPERA, R., STITH, B. J., OLEFSKY, J. M. and KREBA, E. G. (1986). Increased phosphorylation of ribosomal protein S6 following microinjection of insulin receptor-kinase into Xenopus oocytes. Nature 320, 459-461.

MANDEL, M. and HIGA, A. (1970). Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53, 154.

MANIATIS, T., FRITSCH, E. F. and SAMBROOK, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbour, NY.

MARTIN, F. H., CASTRO, M. M., ABOUL-ELA, F. and TINOCO, I. (1985). Base pairing involving deoxyinosine: implications for probe design. Nucl. Acids Res. 13, 8927-8938.

MARTIN-PEREZ, J., RUDKIN, B. B., SIEGMANN, M. and THOMAS, G. (1986). Activation of ribosomal protein S6 phosphorylation during meiotic maturation of Xenopus laevis oocytes: in vitro ordered appearance of S6 phosphopeptides. EMBO 5, 725-731.

MARTIN-PEREZ, J. and THOMAS, G. (1983). Ordered phosphorylation of 40S ribosomal protein S6 after serum stimulation of quiescent 3T3 cells. Proc. Natl. Acad. Sci. USA 80, 926-930.

MATSUDA, Y. and GUROFF, G. (1987). Purification and mechanism of activation of a nerve growth factor-sensitive S6 Kinase from PC12 cells. J. Biol. Chem. 262, 2832-2844.

MATTEUCCI, M. D. and CARUTHERS, M. H. (1981). Nucleotide chemistry, 4. Synthesis of deoxyoligonucleotides on a polymer support. J. Am. Chem. Soc. 103, 3185-3192.

MELTON, D. A., KRIEG, P. A. REBOGLIATI, M. R., MANIATIS, T., ZINN, K. and GREEN, M. R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. 12, 7035-7056.

MESSING, J. (1983). New M13 vectors for cloning. Methods Enzymol. 101, 20-78.

MEYUHAS, O. (1985). Evolutionary conservation of ribosomal protein mRNA sequences: application for expansion of corresponding cDNA and gene libraries. Biochem. and Biophys. Acta, 825, 393-397.

MEYUHAS, O. and PERRY, R. P. (1980). Construction and identification of cDNA clones for mouse ribosomal proteins: application for the study of r-protein gene expression. Gene 10, 113-129.

MEYUHAS, O., THOMPSON, E. A. and PERRY, R. P. (1987). Glucocorticoids selectively inhibit translation of

ribosomal protein mRNAs in P1798 lymphosarcoma cells. Mol. and Cell. Biol. 7, 2691-2699.

MIESFELD, R. and ARNHEIM, N. (1984). Species-specific rDNA transcription is due promoter-specific binding factors. Mol. Cell. Biol. 4, 221-227.

MISHIMA, Y. , FINANCSEK, I., KOMINAMI, R. and MURAMATSU, M. (1982). Fractionation and reconstitution of factors required for accurate transcription of mammalian ribosomal RNA genes: identification of a species-dependent initiation factor. Nucleic Acids Res. 10, 6659-6670.

MONK, R. J., MEYUHAS, O. and PERRY, R.P. (1981). Mammals have multiple genes for individual ribosomal proteins. Cell 24, 301-306.

MULLIGAN, P. K. and HACKETT, P. B. (1984). Isolation of a mouse DNA fragment with homology to a drosophila ribosomal protein gene. GENE 34, 155-161.

MURAMATSU, M., YAMAMOTO, O., KISHIMOTO, T., NAGAMINE, M., KATO, H. and KOMINAMI, R. (1986). Structure and regulation of mammalian ribosomal RNA gene. Adv. Biophys. 21, 217-227.

NAKAMICHIMI, N., RHOADS, D. D. and ROUFA, D. J. (1983). The chinese hamster cell emetine resistance gene. J. Biol. Chem. 258, 13236-13242.

NAKAMICHI, N. N., KAO, F., WASMUTH, J. and ROUFA, D. J. (1986). Ribosomal protein gene sequences map to human chromosomes 5, 8, and 17. Somat. Cell. Genet. 12, 225-236.

NAZAR, R. N., SITZ, T. O. and BUSCH, H.(1975). Homologies in eukaryotic 5.8S ribosomal RNA. Biochem. Biophys. Res. Comm. 62, 736-743.

NOMURA, M., GOURSE, R. and BAUGHMAN, G. (1984). Regulation of the synthesis of ribosomal components. Ann. Rev. Biochem. 53, 75-117.

NOVAK-HOFERT, I. and THOMAS, G. (1984). An activated S6 Kinase in Extracts from Serum- and Epidermal growth factor-stimulated Swiss 3T3 cells. J. Biol. Chem. 259, 5995-6000.

O'BRIEN, S. J. (1987). A compilation of linkage and restriction maps of genetically studied organisms.

Vol 4. Cold Spring Harbour Laboratories, Cold Spring Harbour, N. Y.

OKAYAMA, H. and BERG, P. (1983). A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. Mol. Cell. Biol. 3, 280-289.

OLSON, M. O. J., RIVERS, Z. M., THOMPSON, B. A., KAO, W. Y. and CASE, S. T. (1983). Interaction of nucleolar phosphoprotein C23 with cloned segments of rat ribosomal deoxyribonucleic acid. Biochemistry 22, 3345-3351.

OU, J-H., YEN, T. S. B., WANG, Y-F., KAM, W. K. and RUTTER, W. J. (1987). Cloning and characterization of a human ribosomal protein gene with enhanced expression in fetal and neoplastic cells. Nucl. Acids Res. 15, 8919-8934.

PAGE, D. C., MOSHER, R., SIMPSON, E. M., FISHER, E. M. C., MARDON, G., POLLACK, J., MCGILLVRAY, B., DE LA CHAPELLE, and BROWN, L. G. (1987). The sex-determining region of the human Y chromosome encodes a finger protein. Cell 51, 1091-1104.

PEARSON, N. J., FRIED, H. M. and WARNER, J. R. (1982). Yeast use translation control to compensate for extra copies of a ribosomal protein gene. Cell 29, 347-355.

PELECH, S. L., OLWIN, B. B. and KREBS, E. G. (1986). Fibroblast growth factor treatment of Swiss 3T3 cells activates a subunit S6 kinase that phosphorylates a synthetic peptide substrate. Proc. Natl. Acad. Sci. USA 83, 5968-5972.

PELED-YALIF, E., COHEN-BINDER, I. and MEYUHAS, O. (1984). Isolation and characterization of four mouse ribosomal-protein-L18 genes that appear to be processed psuedogenes. Gene 29, 157-166.

PERISIC, O. and TRAUGH, J. (1983). Protease activated kinase II mediates multiple phosphorylation of ribosomal protein S6 in reticulocytes. J. Biol. Chem. 258, 13998-14002.

PIERANDREI-AMALDI, P., BECARI, E., BOZZONI, I. and AMALDI, F. (1985). Ribosomal protein production in normal and anucleolate Xenopus embryos: regulation at the post transcriptional and translational levels. Cell 42, 317-323.

RHOADS, D. D. and ROUFA, D. J. (1985). Emetine resistance of chinese hamster cells: Structures of wild type and mutant ribosomal protein S14 mRNAs. Mol. Cell. Biol. 5, 1655-1659.

RHOADS, D. D. and ROUFA, D. J. (1987). A cloned human ribosomal protein gene functions in rodent cells. Mol. and Cell. Biol. 7, 3767-3774.

RHOADS, D. D., DIXIT, A. and ROUFA, D. J. (1986). Primary structure of human ribosomal protein S14 and the gene that encodes it. Mol. Cell. Biol. 6, 2774-2783.

RIGBY, P. W., DIEKMANN, M., RHODES, C., and BERG, P. (1977). Labelling of deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase. J. Mol. Biol. 113, 237-251.

ROTENBERG, M. O. and WOOLFORD, J. L. (1986). Tripartite upstream promoter element essential for expression of Saccharomyces cerevisiae ribosomal protein genes. Mol. and Cell. Biol. 6, 674-687.

SAKONJU, S., BOGENHAGEN, D. F. and BROWN, D. D. (1980). A control region in the centre of the 5S RNA gene directs specific initiation of transcription in the 5' border of the region. Cell 19, 13-25.

SAKONJU, S. and BROWN, D. D. (1982). Contact points between a positive transcription factor and the Xenopus 5S RNA gene. Cell 31, 395-405.

SAMUELS, M., FIRE, A. and SHARP, P. A. (1984). Dinucleotide priming of transcription mediated by RNA polymerase II. J. Biol. Chem. 259, 2517-2525.

SCHEER, U., LANFRANCHI, G., ROSE, K. M., FRANKE, W. W. and RINGERTZ, N. R. (1983). Migration of rat RNA polymerase I into chick erythrocyte nuclei undergoing reactivation in chick-rat heterokaryons. J. Cell Biol. 97, 1641-1643.

SCHEER, U. and ROSE, K. M. (1984). Localization of RNA polymerase I in interphase and mitotic chromosomes by light and electron microscope immunocytochemistry. Proc. Natl. Acad. Sci. USA 81, 1431-1435.

SCHMICKEL, R. D. (1973). Quantitation of human ribosomal DNA: Hybridization of human DNA with ribosomal RNA for quantitation and fractionation. Pediat. Res. 7, 5-12.

SCHMIDT-ZACHMANN, M. S., HUGLE, B., SCHEER, U. and FRANKE, W. (1984). Identification and localization of a novel nucleolar protein of high molecular weight by a monoclonal antibody. Experimental Cell Res. 153, 327-346.

SEGALL, J., MATSUI, T. and ROEDER, R. G. (1980). Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. J. Biol. Chem. 255, 11986-11991.

SOMMERVILLE, J. (1985). Organizing the nucleolus. Nature 318, 410-411.

SPECTOR, D. L., OCHS, R. L. and BUSCH, H. (1984). Silver staining, immunofluorescence, and immunoelectron microscopic localization of nucleolar phosphoproteins B23 and C23. Chromosoma 90, 139-148.

STEFFENSEN, D. M., DUFFY, P. and PRENSKY, W. (1974). Localization of 5S rRNA genes on human chromosome I. Nature 252, 741-743.

STROHMAN, R. C., MOSS, P. S., MICOU-EASTWOOD, J. and SPECTOR, D. (1977). Messenger RNA for myosin polypeptides: isolation from single myogenic cell cultures. Cell 10, 265-273.

SUGGS, S., WALLACE, R. B., HIROSE, T., KAWSHIMA, E. H. and ITAKURA, K. (1981). Use of synthetic oligonucleotides as hybridization probes: Isolation of cloned cDNA for human beta 2-microglobulin. Proc. Natl. Acad. Sci. USA 78, 6613-6617.

TANAKA, T. KUWANO, Y., ISIKAWA, K. and OGATA, K. Nucleotide sequence of cloned cDNA specific for rat ribosomal protein S11. J. Biol. Chem. 260, 6329-6333.

TANAKA, T., WOOL, I. G., STOFFLER, G., (1980). The effect of antibodies against Escherichia coli small ribosomal subunit proteins on protein synthesis by rat liver ribosomes. J. Biol. Chem. 255, 3832-3834.

TODOROV, I. T., NOLL, F. and HADJIOLOV, A. A. (1983). The sequential addition of ribosomal proteins during the formation of the small ribosomal subunit in Friend erythroleukemia cells. Eur. J. Biochem. 131, 271-275.

TOLAN, D. R. and TRAUT, R. R. (1981). Protein topography of the 40 S ribosomal subunit from rabbit reticulocytes shown by cross-linking with 2-iminothiolane. J. Biol. Chem. 256, 10129-10136.

THOMAS, G., SIEGMANN, M., KUBLER, A. M., GORDON, J. and JIMENEZ DE ASUA, L. (1980). Regulation of 40S ribosomal protein S6 phosphorylation in Swiss mouse 3T3 cells. Cell 19, 1015-1023.

THOMAS, G., MARTIN-PEREZ, J., SIEGMANN, M. and OTTO, A. M. (1982). The effect of serum, EGF, PGF2 beta and insulin on S6 phosphorylation and the initiation of protein and DNA synthesis. Cell 30, 235-242.

ULLRICH, A., BERMAN, C. H., DULL, T. J., GRAY, A. and LEE, J. M. (1984). Isolation of the human insulin growth factor 1 gene using a single synthetic DNA probe. Embo J. 3, 361-364.

ULLRICH, A., COUSSENS, L., HAYFLICK, J. S., DULL, T. J., GRAY, A., TAM, A. W., LEE, J., YARDEN, Y., LIBERMAN, T. A., SCHLESSINGER, J., DOWNARD, J., MAYES, E. L. V., WHITTLE, N., WATERFIELD, M. D. and SEEBURG, P. H. (1984). Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature 309, 418-425.

VASLET, C. A., O'CONNELL, P., IZQUIERDO, M. and ROSBASH, M. (1980). Isolation and mapping of a cloned ribosomal protein gene of drosophila melanogaster. Nature 285, 674-676.

WAGNER, M. and PERRY, R. P. (1985). Characterization of the multigene family encoding the mouse S16 ribosomal protein: strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. Mol. and Cell. Biol. 5, 3560-3576.

WARNER, J. R. (1979). Distribution of newly formed ribosomal proteins in HeLa cells fractions. J. Cell. Biol. 80, 767-772.

WARNER, J. R. and GORENSTEIN, C. (1979). The ribosomal proteins of Saccharomyces cerevisiae. Meth. Cell. Biol. 20, G. Stein, J. Stein and L. J. Kleinsmith, (New York Acad. Press), 45-60.

WARNER, J. R., MITRA, G., SCHWINDER, W. F., STUDNEY, M. and FRIED, H. M. (1985). Saccharomyces cerevisiae

coordinates accumulation of yeast ribosomal proteins by modulating mRNA splicing, translational iniation, and protein turnover. Mol. and Cell. Biol. 5, 1512-1521.

WARNER, J. R. and SOEIRO, R. (1967). Nascent ribosomes from HeLa cells. Proc. Natl. Acad. Sci. USA 58, 1981-1990.

WELLAUER, P. K. and DAWID, I. B. (1973). Secondary structure maps of RNA. Processing of HeLa rRNA. Proc Natl. Acad. Sci. USA 70, 2827-2831.

WETTENHALL, R. E. H. and MORGAN, F. J. (1984). Phosphorylation of hepatic ribosomal protein S6 on 80 and 40 S ribosomes. J. Biol. Chem. 259, 2084-2091.

WIEDEMANN, L. M. and PERRY, R. P. (1984). Characterization of the expressed gene and several processed psuedogenes for the mouse ribosomal protein L30 gene family. Mol. Cell. Biol. 4, 2518-2528.

WOOL, I. G.(1979). The structure and function of eukaryotic ribosomes. Ann. Rev. Biochem. 48, 719-754.

WOOLFORD, J. L., HEREFORD, L. M. and ROSBASH, M. (1979). Isolation of cloned DNA sequences containing ribosomal protein genes from Saccharomyces cerevisiae. Cell 18, 1247-1259.

WORKMAN, J. L. and ROEDER, R. G. (1987). Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. Cell 51, 613-622.

YOUNG, J. A. T., and TROWSDALE, J. (1985). A processed pseudogene in an intron of the HLA-DP beta 1 chain is a member of the ribosomal protein L32 gene family. Nucleic Acids Res. 13, 8883-8891.

YOUNG, R. A. and DAVIS, R. W. (1983). Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA 80, 1194-1198.

ZINKER, S and WARNER, J. R. (1976). The ribosomal proteins of Saccharmyces cerevisiae. J. Biol. Chem. 251, 1799-1807.

ZUKER, C. S., COWMAN, A. F. and RUBIN, G. M. (1985). Isolation and structure of a rhodopsin gene from D. melanogaster. Cell 40, 851-858.